

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2003, 15:03:04 ; Search time 29 Seconds  
(without alignments)  
77.327 Million cell updates/sec

Title: US-10-016-768a-2

Perfect score: 273  
Sequence: 1 KQRRKKRGRRYDHEIMEE.....KAQIGVPHSTLEYKVER 53

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	59	21.6	191	2	US-08-286-819A-21 Sequence 21, Appl
2	59	21.6	191	3	US-08-980-357-21 Sequence 21, Appl
3	57	20.9	102	4	US-09-732-210-1293 Sequence 1293, Ap
4	56	20.5	696	4	US-09-107-532A-6133 Sequence 6133, Ap
5	55.5	20.3	104	4	US-09-732-210-1272 Sequence 1272, Ap
6	54	19.8	625	4	US-08-252-991A-33087 Sequence 33087, A
7	53.5	19.6	187	4	US-09-328-352-5117 Sequence 5117, A
8	53.5	19.6	438	4	US-09-134-001C-3634 Sequence 3634, Ap
9	53	19.4	137	4	US-09-107-532A-4469 Sequence 4469, Ap
10	53	19.4	227	4	US-09-308-003-48 Sequence 48, Appl
11	53	19.4	390	4	US-09-308-003-12 Sequence 12, Appl
12	52.5	19.2	389	4	US-09-252-991A-24471 Sequence 24471, A
13	52.5	19.2	810	4	US-09-540-824-25 Sequence 25, Appl
14	52	19.0	552	4	US-08-974-546-1 Sequence 1, Appl
15	52	19.0	552	4	US-09-446-301A-4 Sequence 4, Appl
16	52	19.0	552	4	US-09-099-932-4 Sequence 4, Appl
17	52	19.0	560	4	US-09-446-301A-50 Sequence 50, Appl
18	52	19.0	663	1	US-07-912-015-2 Sequence 2, Appl
19	51.5	18.9	349	3	US-08-462-467B-18 Sequence 18, Appl
20	51.5	18.9	554	3	US-08-462-467B-22 Sequence 22, Appl
21	51.5	18.9	1399	3	US-08-462-467B-14 Sequence 14, Appl
22	51	18.7	101	4	US-09-732-210-1268 Sequence 1268, Ap
23	51	18.7	218	4	US-09-107-532A-5919 Sequence 5919, Ap
24	51	18.7	222	4	US-09-252-991A-20670 Sequence 20670, A
25	51	18.7	425	3	US-08-462-467B-16 Sequence 16, Appl
26	51	18.7	649	3	US-08-462-467B-20 Sequence 20, Appl
27	51	18.7	1495	3	US-08-462-467B-12 Sequence 12, Appl

28	50.5	18.5	113	4	US-09-107-532A-6213 Sequence 6213, Ap
29	50.5	18.5	521	4	US-09-252-991A-30932 Sequence 30932, A
30	50	18.3	60	3	US-08-928-213B-184 Sequence 184, Ap
31	50	18.3	100	2	US-08-160-524A-12 Sequence 12, Appl
32	50	18.3	278	3	US-08-821-278A-18 Sequence 18, Appl
33	50	18.3	346	2	US-08-613-220B-4 Sequence 4, Appl1
34	50	18.3	398	4	US-09-252-991A-21127 Sequence 21127, A
35	50	18.3	519	4	US-09-107-532A-7161 Sequence 7161, Ap
36	50	18.3	802	3	US-09-081-345-18 Sequence 18, Appl
37	49.5	18.1	697	4	US-09-252-991A-27283 Sequence 27283, A
38	49	17.9	1209	4	US-08-252-991A-25844 Sequence 25844, A
39	49	17.9	2584	3	US-08-936-135-4 Sequence 4, Appl
40	48.5	17.8	102	4	US-09-732-210-1283 Sequence 1283, Ap
41	48.5	17.8	104	4	US-09-732-210-144 Sequence 144, App
42	48.5	17.8	352	4	US-09-996-243-505 Sequence 505, App
43	48.5	17.8	361	4	US-09-198-452A-812 Sequence 812, App
44	48.5	17.8	365	2	US-08-979-424-3 Sequence 3, Appl1
45	48.5	17.8	365	3	US-08-928-383B-2 Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-286-819A-21  
: Sequence 21, Application US/08286819A  
: Patent No. 5871910  
: GENERAL INFORMATION:  
: APPLICANT: ARTHUR, MICHEL  
: APPLICANT: DUKTA-MALEN, SYLVIE  
: APPLICANT: MOLINAS, CATHERINE  
: APPLICANT: COURVALIN, PATRICE  
: TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
: TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERES, IN PARTICULAR  
: TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
: NUMBER OF SEQUENCES: 54  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
: ADDRESSEE: P.C.  
: STREET: 1755 S. Jefferson Davis Highway, Suite 400  
: CITY: Arlington  
: STATE: Virginia  
: COUNTRY: U.S.A.  
: ZIP: 22202  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/286.819A  
: FILING DATE: 05-AUG-1994  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/174,682  
: FILING DATE: 28-DEC-1993  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/917,146  
: FILING DATE: 10-AUG-1992  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/FR/91/00855  
: FILING DATE: 29-OCT-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: FR 9013579  
: FILING DATE: 31-OCT-1990  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Oblon, No. 5871910man F.  
: REGISTRATION NUMBER: 24,618  
: REFERENCE/DOCKET NUMBER: 660-060-0 PCT

```

      TELECOMMUNICATION INFORMATION:  
        TELEPHONE: (703) 413-3000  
        TELEFAX: (703) 413-2220  
        TELEX: 248855 OPAT UR  
    INFORMATION FOR SEQ ID NO: 21:  
        SEQUENCE CHARACTERISTICS:  
          LENGTH: 191 amino acids  
          TYPE: amino acid  
          TOPOLOGY: linear  
        MOLECULE TYPE: protein  
US-08-286-819A-21
```

Query Match                  21.6%; Score 59; DB 2; Length 191;  
Best Local Similarity      29.6%; Pred. No. 0.71;  
Matches    16; Conservative     13; Mismatches    23; Indels        2; Gaps        1;

Oy                  1 KQPKKAGRRQR-DHEIMEEAIAMVSGSKSVSAQCIGVPHPSTLEYKYKE 52  
         | : || :| : | : | : | : | : | : | :  
Db                  135 KKGEKFGRLLKKYKNHAGMNVAVKLKGGMNTVNQICEITTVSRASLYRKLE 188

RESULT 2  
Sequence 21, Application US/08980357  
Patent No. 6013508  
GENERAL INFORMATION:  
APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKRA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR  
IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
NUMBER OF INVENTIONS: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSER: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
City: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/980,357  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,819  
FILING DATE: 05-AUG-1994  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9013579  
FILING DATE: 31-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 6013508man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220

```

TELEX: 248655 OPAT UR
INFORMATION FOR SEQ ID NO.: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-980-357-21

Query Match      21.6%; Score 59; DB 3; Length 191;
Best Local Similarity 29.6%; Pred. No. 0.71;
Matches 16; Conservative 13; Mismatches 23; Indels 2; Gaps 1;

QY      1 KKKRRKRYQY--DHEIMEEAIAMWMSGMSYSKAAGIYGVSHTLEKYKE 52
       |:|::||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      135 KKGFGRKLKRYKHKNHAGMNYAVKLYKEGMMVTNOICETITNYSRASYRLSE 188

RESULT 3
US-09-732-210-1293
Sequence 1293, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunkers, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Miltanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-fungal Proteins and Methods for their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 1293
LENGTH: 102
TYPE: PRT
ORGANISM: Thermotoga maritima
US-09-732-210-1293

Query Match      20.9%; Score 57; DB 4; Length 102;
Best Local Similarity 26.8%; Pred. No. 0.62;
Matches 15; Conservative 14; Mismatches 17; Indels 10; Gaps 2;

QY      4 RKKRRRYQYQVHEIMEEAIAMWMS-GKRSYSKAG-----IYGVPHSTLEYK 49
       |:|:|:||||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      4 OKIRIKLAYDEHIDESAKRIIVEASTNSKSGPIPLPTERTLYCLVLSPMKHK 59

RESULT 4
US-09-107-532A-6133
Sequence 6133, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Maitland
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

```



```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/107,532A
: FILING DATE: 30-Jun-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/085,598
: FILING DATE: 14 May 1998
: APPLICATION NUMBER: 60/051571
: FILING DATE: July 2, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Arinello, Pamela Deneko
: REGISTRATION NUMBER: 40,489
: TELECOMMUNICATION INFORMATION:
: REFERENCE/DOCKET NUMBER: GTC-012
: TELEPHONE: (781)893-5007
: TELEFAX: (781)893-8277
: INFORMATION FOR SEQ ID NO: 6133:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 696 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (B) LOCATION 1...696
: SEQUENCE DESCRIPTION: SEQ ID NO: 6133:
US-09-107-532A-6133

Query Match      20.5% Score 56; DB 4; Length 696;
Best Local Similarity 32.1% Pred. No. 11;
Matches 18; Conservative 8; Mismatches 18; Indels 12; Gaps 3;

QY      3 PRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKV-----KER 53
DB      372 PTEENRAREYKLSNE--ERKIV-----OMIVSRFLGLAOPHKVSOFTVVEFGKER 420

RESULT 5
US-09-732-210-1272
: Sequence 1272, Application US/09732210
: Patent No. 6573361
: GENERAL INFORMATION:
: APPLICANT: Bunkers, Greg J.
: APPLICANT: Liang, Jihong
: APPLICANT: Mittanck, Cindy A.
: APPLICANT: Seale, Jeffrey W.
: APPLICANT: Wu, Yonnie S.
: TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
: FILE REFERENCE: 38-21(15036)B
: CURRENT APPLICATION NUMBER: US/09/732,210
: CURRENT FILING DATE: 2000-12-07
: PRIOR APPLICATION NUMBER: US 60/169,513
: PRIOR FILING DATE: 1999-12-07
: PRIOR APPLICATION NUMBER: US 60/169,340
: PRIOR FILING DATE: 1999-12-07
: NUMBER OF SEQ ID NOS: 1753
: SEQ ID NO 1272
: LENGTH: 104
: TYPE: PRT
: ORGANISM: Helicobacter pylori
US-09-732-210-1272

Query Match      20.3% Score 55.5; DB 4; Length 104;
Best Local Similarity 29.8% Pred. No. 1.1;
Matches 14; Conservative 12; Mismatches 20; Indels 1; Gaps 1;

QY      5 KKRGYROYDHEIMEEA-IAVMGSKMSVSKAGIYGVPHSTLEYKV 50
DB      3 KIRKLKAYDHRVLDRSVAIVAEVAKRSGSEIRPIPLPKNKRYTV 49
```

```

:
: RESULT 6
: US-09-252-991A-33087
: Sequence 33087, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196,136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 33087
: LENGTH: 625
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33087

Query Match      19.8% Score 54; DB 4; Length 625;
Best Local Similarity 36.1% Pred. No. 18;
Matches 13; Conservative 7; Mismatches 12; Indels 4; Gaps 1;

QY      6 KRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKV 37
DB      90 ERGHRRLGIAFQHEVAEAREAPVGLAEQLAEG 125

RESULT 7
US-09-328-352-5117
: Sequence 5117, Application US/09328352
: Patent No. 6562958
: GENERAL INFORMATION:
: APPLICANT: Gary L. Breton et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
: FILE REFERENCE: GTC99-03BA
: CURRENT APPLICATION NUMBER: US/09/328,352
: CURRENT FILING DATE: 1999-06-04
: NUMBER OF SEQ ID NOS: 8252
: SEQ ID NO 5117
: LENGTH: 187
: TYPE: PRT
: ORGANISM: Acinetobacter baumannii
US-09-328-352-5117

Query Match      19.6% Score 53.5; DB 4; Length 187;
Best Local Similarity 31.2% Pred. No. 4.5;
Matches 15; Conservative 12; Mismatches 16; Indels 5; Gaps 2;

QY      10 YROYDHEIMEAIAVMGSKMSV-SKAQIY---GVPHSTLEYKVE 52
DB      64 YQARNYEIKIEVINLFLKISDSIHKLKGLYDGLTKNSDEPKYKE 111

RESULT 8
US-09-134-001C-3634
: Sequence 3634, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
```



```

: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 24471
: LENGTH: 389
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24471

Query Match 19.2%; Score 52.5; DB 4; Length 389;
Best Local Similarity 22.8%; Pred. No. 16;
Matches 13; Conservative 15; Mismatches 16; Indels 13; Gaps 2;

QY 8 GRVROYHEIMEAIA-----MVMSGKMSVSKAGIYGVPHSTLEKVKER 53
DB 217 GEFOGAHEADRAVAGQRLLEVEDVYVAGGAGIASDQHY--PHSRIGLRLOQR 271.

RESULT 13
US-09-540-824-25
: Sequence 25, Application US/09540824
: Patent No. 6383753
: GENERAL INFORMATION:
: APPLICANT: Thiele, Dennis
: TITLE OF INVENTION: No. 6383753el Yeast and Mammalian Regulators of Cell Proliferat
: FILE REFERENCE: UM-04286
: CURRENT APPLICATION NUMBER: US/09/540,824
: CURRENT FILING DATE: 2000-03-31
: NUMBER OF SEQ ID NOS: 28
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 25
: LENGTH: 810
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-540-824-25

Query Match 19.2%; Score 52.5; DB 4; Length 810;
Best Local Similarity 36.8%; Pred. No. 42;
Matches 14; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

QY 1 KOPRRKRGYROYDHEIMEAIAVMSGKMSVSKAGI 38
DB 199 QOPORRKT-----AEVWKEVIASKPYKQEROKRQGI 231

RESULT 14
US-08-974-546-1
: Sequence 1, Application US/08974546
: Patent No. 5945287
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
```

```

: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/974,546
: FILING DATE: Filed Herewith
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0428
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 348 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: BRAINTRUT21
: CLONE: 2525691
US-08-974-546-1

Query Match 19.0%; Score 52; DB 2; Length 348;
Best Local Similarity 39.0%; Pred. No. 17;
Matches 16; Conservative 4; Mismatches 19; Indels 2; Gaps 1;
```

```
QY 5 KRGYROYDHEIMEAIAVMSGKMSVSKAGIYGVPHST 45
DB 59 KRGGLDYDGEGLKTGGC--TSGGSSGSFHYTFHGDPHAT 97
```

```

RESULT 15
US-09-446-301A-4
: Sequence 4, Application US/09446301A
: Patent No. 6506893
: GENERAL INFORMATION:
: APPLICANT: EL SOLH, NEVINE
: APPLICANT: ALIENNET, JEANINE
: TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
: FILE REFERENCE: 03715-0059
: CURRENT APPLICATION NUMBER: US/09/446,301A
: CURRENT FILING DATE: 1999-12-20
: NUMBER OF SEQ ID NOS: 51
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 4
: LENGTH: 552
: TYPE: PRT
: ORGANISM: Staphylococcus sp.
US-09-446-301A-4

Query Match 19.0%; Score 52; DB 4; Length 552;
Best Local Similarity 31.1%; Pred. No. 30;
Matches 14; Conservative 11; Mismatches 12; Indels 8; Gaps 2;
```

```
QY 4 RKRGRYROYD-HEIMEAIAVMSGKMSVSKAGIYGVPHSTL 46
DB 193 RKQDEYEKYSKRKOLEAI-----KLKENKAGMIRPPSKTM 231
```

Search completed: August 4, 2003, 15:08:06  
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2003, 14:55:14 ; Search time 26 seconds  
(without alignments)  
95.862 Million cell updates/sec

Title: US-10-016-768a-2

Perfect score: 273  
Sequence: 1 KOPRRKKRGYRQYDHEIMEE.....KAQIGYVPHSTLEYKVKER 53

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.5	25.5	753	1	TKR_DROME
2	63	23.1	977	1	BAB1_DROME
3	61	22.3	352	1	VFDE_ECOLI
4	59	21.6	191	1	TNR6_ENTFC
5	59	21.6	1067	1	BAB2_DROME
6	58	21.2	342	1	REC1_MYXXA
7	58	21.2	397	1	CISR_STYX3
8	57	20.9	102	1	RS10_THEMA
9	56.5	20.5	102	1	RS10_THEMA
10	56	20.5	102	1	RS10_LISMO
11	55.5	20.3	104	1	RS10_HELPY
12	55.5	20.3	104	1	RS10_HELPY
13	55.5	20.3	169	1	VY69_METJA
14	55	20.1	236	1	VY46_CHLTR
15	55	20.1	420	1	SYH_STRAM
16	55	20.1	765	1	PURL_ARCFU
17	55	20.1	809	1	QUTR_ACICA
18	55	20.1	882	1	ALAS_PSEOL
19	54.5	20.0	2607	1	BACB_BACLI
20	54	19.8	352	1	RECA_BORPE
21	54	19.8	764	1	METE_SCHPO
22	54	19.8	936	1	FHUI_YEAST
23	54	19.8	1220	1	DP3A_THEAO
24	54	19.8	1547	1	TOR2_BOMO
25	53.5	19.6	138	1	VFEO_ECOLI
26	53.5	19.6	258	1	BPHF_ECOLI
27	53.5	19.6	375	1	V785_METJA
28	53.5	19.6	714	1	PURL_METJA
29	53	19.4	102	1	RS10_BACHD
30	53	19.4	102	1	RS10_STRCO
31	52.5	19.2	344	1	CHEB_CAUCR
32	52.5	19.2	478	1	FXD3_HUMAN
33	52.5	19.2	1390	1	N155_RAT

34	52.5	19.2	1391	1	N155_MOUSE
35	52	19.0	102	1	RS10_STARN
36	52	19.0	102	1	RS10_STARN
37	52	19.0	264	1	TRA2_DROME
38	52	19.0	348	1	DJB5_HUMAN
39	52	19.0	715	1	PURL_METAC
40	52	19.0	964	1	UL70_MCVMS
41	51.5	18.9	417	1	EX7L_LACIA
42	51.5	18.9	900	1	SYA_AERPE
43	51.5	18.9	1391	1	N155_HUMAN
44	51	18.7	101	1	RS10_BACSU
45	51	18.7	101	1	RS10_CORGL

## ALIGNMENTS

RESULT 1	TKR_DROME	STANDARD:	PRT:	753 AA.
ID	TKR_DROME			
AC	P14083:			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Protein TKR.			
GN	TKR.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila..			
OX	NCBI_Taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=8112827; PubMed=3428600;			
RA	Haller J., Cole S., Broenner G., Jaeckle H.;			
RT	"Dorsal and neural expression of a tyrosine kinase-related Drosophila			
RT	gene during embryonic development.";			
RL	Genes Dev. 1:862-867(1987).			
CC	- FUNCTION: POSSIBLE REGULATORY ROLE DURING DEVELOPMENT.			
CC	- CAUTION: WAS ORIGINALLY THOUGHT TO BE A KINASE ON THE BASIS OF			
CC	WEAK AND NON-SIGNIFICANT SIMILARITIES.			
CC	PIR: A27041; A27041.			
DR	FLYBase: FBgn0003715; TKR.			
DR	Pfam: PF05225; HTH_psq: 1.			
FT	DOMAIN	143	151	POLY-ASP.
FT	DOMAIN	153	157	POLY-GLU.
FT	DOMAIN	174	183	POLY-ALA.
FT	DOMAIN	221	224	POLY-ASN.
FT	DOMAIN	297	306	POLY-ALA.
FT	DOMAIN	325	332	POLY-ALA.
FT	DOMAIN	709	712	POLY-ALA.
SQ	SEQUENCE	753 AA;	81021 MW;	P98D3272A7DDBESE CRC64;
Query Match	25.5%	Score 69.5;	DB 1;	Length 753;
Best local Similarity	32.1%;	Pred. No. 0.31;		
Matches	17;	Conservative 12;	Mismatches 17;	Indels 7;
			Gaps 1;	
QY	1 KOPRRK-----RGRAYDYDHEIMEEAIAMVMSGKMSVKAQIGYVPHSTL 46			
DB	496 KMRPRKQGHAPRGPGPRSWTNTLFTALQHVNNKKMTTSQASRIFETIPYNSL 548			
RESULT 2	BAB1_DROME	STANDARD:	PRT:	977 AA.
ID	BAB1_DROME			
AC	G9W0K7: Q23968; Q8WR78; Q9U1H7;			
DT	15-SEP-2003 (Rel. 42, Created)			
DT	15-SEP-2003 (Rel. 42, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Bric-a-brac protein 1.			
GN	BAB1 OR BAB OR CG9097/CG13910.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			



```

Query Match          23.1%; Score 63; DB 1; Length 977;
Best Local Similarity 34.1%; Pred. No. 3.1;
Matches    14; Conservative   10; Mismatches    17; Indels      0; Gaps      0;

Oy      5 KKRGRYROYDHEIMEEAIAMWSGKMSYSKAQGIYGVPHST 45
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      560 RERGPLKSMREPMETMAEAIFSLVKGSLGSQAARKARYDIPPT 600

RESULT 3
VFDF_ECOLI STANDARD: PRT: 352 AA.
AC P76505;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DS Hypothetical protein yfde.
GN VFDF OR B2345.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC DR EMBL; AE000323; AAC75405.1; .
DR PIR; G65007; G65007.
DR EcGene; EG14130; yfde.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 352 AA; 40248 MW; 632F3D31C139491 CRC64;

Query Match          22.3%; Score 61; DB 1; Length 352;
Best Local Similarity 28.8%; Pred. No. 1.7;
Matches    17; Conservative   15; Mismatches    15; Indels     12; Gaps      3;

Oy      5 KKRGRYROYDIET-----MEEAIAMWSGKMSYSKAQGIYGVPH-STLETKYKER 53
       |::|::|::|::|::|::|::|::|::|::|::|::|
Db      271 KRVGRAIRYDFOVDAKYKCDHLEIKETIVSTINGKNINWGHOSKIY-KHKFDLGEKIEER 327

RESULT 4
TNR6_ENTFC STANDARD: PRT: 191 AA.
AC Q06237;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DS Transposon Tn1546 resolvase.
OS Enterococcus faecium (Streptococcus faecium).
OC Plasmid pJP816.
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1352;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=B4147;
RX MEDLINE=93106944; PubMed=8380148;

```

```

RA Arthur M., Molinas C., Depardieu F., Courvaulin P.:
RT "Characterization of Tn1546, a Tn3-related transposon conferring
RT glycopeptide resistance by synthesis of depeptide peptidoglycan
RT precursors in Enterococcus faecium BM4147."
RL J. Bacteriol. 175:117-127(1993).
CC -1- FUNCTION: RESOLVASE CATALYZES THE RESOLUTION (A SITE-SPECIFIC
CC RECOMBINATION) OF THE INTEGRATED REPLICON TO YIELD THE FINAL
CC TRANSPOSITION PRODUCTS.
CC -1- SIMILARITY: BELONGS TO THE SITE-SPECIFIC RECOMBINASE RESOLVASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M97297; AAA65952.1; -.
DR PIR; B40628; B40628.
DR HSSP; P03012; 2RSL.
DR InterPro; IPR006120; HTH_7.
DR InterPro; IPR006118; Recombinase.
DR InterPro; IPR006119; resolvase_N.
DR Pfam; PF02796; HTH_7; 1.
DR Pfam; PF02939; resolvase; 1.
DR PROSITE; PS00397; RECOMBINASES_1; 1.
DR PROSITE; PS00398; RECOMBINASES_2; FALSE_NEG.
KW DNA recombination; DNA integration; DNA-binding; Transposable element;
KW Plasmid.
FT ACT_SITE 10 10 TRANSIENT COVALENT LINKAGE TO DNA DURING
FT FT STRAND CLEAVAGE AND REJOINING
FT FT (BY SIMILARITY).
SQ SEQUENCE 191 AA; 22296 MW; B027350864EFC17C CRC64;
Query Match 21.6%; Score 59; DB 1; Length 191;
Best Local Similarity 29.6%; Pred. NO. 1.5;
Matches 16; Conservative 13; Mismatches 23; Indels 2; Gaps 1;
QY 1 KQPRKKRRRYQY--DHEIMEEATAMWSGKMSKAGIYGPHSTLEYKKE 52
Db 135 KKEGFKGRLLKKYKHNAGMAYAKLYKEGMVTNVCETITVSRASLYRLE 188
RESULT 5
BAB2_DROME STANDARD: PRT; 1067 AA.
ID BAB2_DROME
AC Q9W0K4; Q24001; Q9U1H3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bric-a-brac protein 2.
DE BAB2 OR FBBI1 OR CG9102/CG13911.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empyridioidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Embryo, and Ovary;
RX MEDLINE=21969340; PubMed=11973274;
RA Couderc J.L.G., Godt D., Zollman S., Chen J., Li M., Tjong S.,
RA Cramton S.E., Saun-Barnoia I., Laski F.A.;
RA "The bric a brac locus consists of two paralogous genes encoding
RT B7B/P20 domain proteins and acts as a homeotic and morphogenetic
RT regulator of imaginal development in Drosophila."
RL Development 129:2419-2433(2002).
RP [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN-Berkeley;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Motman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Foster C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sliden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 [3]  
 RP REVISIONS.  
 RC STRAIN-Berkeley;  
 RX MEDLINE-22426069; PubMed-12531572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.B., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP SEQUENCE OF 196-310 FROM N.A.  
 RX MEDLINE-95024186; PubMed-7938017;  
 RA Zollman S., Godt D., Prive G.G., Couderc J.L., Laski F.A.;  
 RT "The BTB domain, found primarily in zinc finger proteins, defines an  
 RT evolutionarily conserved family that includes several developmentally  
 RT regulated genes in *Drosophila*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:10717-10721(1994).  
 CC -1- FUNCTION: Probably acts as a transcriptional regulator. Required  
 CC for the specification of the tarsal segment. Also involved in  
 CC antenna development.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Leg imaginal disk at the central region of the  
 CC tarsus and in eye antenna disk at the basal cylinder.  
 CC -1- MISCELLANEOUS: 'birc-a-brac' means 'jumble' in French (referring to  
 CC the mutant ovary phenotype).  
 CC -1- SIMILARITY: Contains 1 A-T hook DNA-binding repeat.  
 CC -1- SIMILARITY: Contains 1 BTB/POZ domain.  
 CC -1- SIMILARITY: Contains 1 helix-turn-helix Pqg-type domain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ252173; CAB64388.1; -  
 CC EMBL: AE003470; AAF47442.2; -  
 CC EMBL: U14399; AAB50834.1; -  
 CC FlyBase: FBgn0025525; bab2.  
 CC InterPro: IPR000637; AT hook.  
 CC InterPro: IPR000210; BTB\_POZ.  
 CC InterPro: IPR007889; HTH\_Psq.  
 CC Pfam: PF00651; BTB; 1.  
 CC Pfam: PF05225; HTH\_Psq; 1.  
 CC SMART: SM00225; BTB; 1.  
 CC PROSITE: PS50097; BTB; 1.  
 CC Nucleic acid binding: DNA-binding; Transcription regulation.  
 CC FT DOMAIN 223 288 BTB.  
 CC FT DNA\_BIND 645 690 H-T-H MOTIF PSQ-TYPE.  
 CC FT DNA\_BIND 697 708 A-T HOOK.  
 CC FT CONFLICT 858 858 A -> R (IN REF. 1).  
 CC SO SEQUENCE 1067 AA; 114661 MW; 7DBFC7681D507FC0 CRC64;  
 CC -----  
 CC Query Match 21.6%; Score 59; DB 1; Length 1067;  
 CC Best Local Similarity 31.7%; Pred. No. 12;  
 CC Matches 13; Conservative 11; Mismatches 17; Indels 0; Gaps 0;  
 CC -----  
 CC 5 KKRGRYQYDHEIMEEIAIWMSCMSKVSRAOGIYGVPHST 45  
 CC Db 636 RERGPLKSMRPEAAEAIFSVKLGSLSSQARAFDIPFT 676  
 CC -----  
 CC RESULT 6  
 CC RECD MYXXA STANDARD; PRT; 342 AA.  
 CC ID P48291;  
 CC AC 01-FEB-1996 (Rel. 33, Created)  
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE RecA protein 1 (Recombinase A 1).  
 CC GN RECA1.  
 CC OS Myxococcus xanthus.  
 CC CC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 CC CC Cytophacteryae; Myxococcaceae; Myxococcus.  
 CC OX NCBI\_TaxID=34;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE-95332233; PubMed-7608099;  
 RA Nottola N., Hsu M.-Y., Inouye S., Inouye M.;  
 RT "Two recA genes in *Myxococcus xanthus*.";  
 RL J. Bacteriol. 177:4179-4182(1995).  
 CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF  
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED  
 CC DNA BY DUPLEX DNA. AND THE ATP-DEPENDENT HYBRIDIZATION OF  
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING  
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR RECA IN M.XANTHUS; RECA1  
 CC SEEMS LESS FUNCTIONAL THAN RECA2.  
 CC -1- SIMILARITY: Belongs to the recA family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



DR EMBL: LA0367: AAC37000.1: -  
 DR PIR: A57364: A57364.  
 DR HSSP: P26345: 1G19.  
 DR HAMAP: MF\_00268: 1.  
 DR InterPro: IPR003593: AAA\_ATPase.  
 DR InterPro: IPR001553: RECA.  
 DR Pfam: PF00154: RECA.1.  
 DR PRINTS: PR00142: RECA.  
 DR ProDom: PD000229: RECA.1.  
 DR SMART: SM00382: AAA.1.  
 DR PROSITE: PS00321: RECA\_1; 1.  
 DR PROSITE: PS50162: RECA\_2; 1.  
 DR PROSITE: PS50163: RECA\_3; 1.  
 DR DNA damage: DNA recombination: SOS response; ATP-binding; DNA-binding.  
 FT NP\_BIND 68 75 ATP (BY SIMILARITY)  
 SQ SEQUENCE 342 AA; 36055 MW; 7F868EAF9FB47E4 CRC64;

Query Match 21.2%; Score 58; DB 1; Length 342;  
 Best Local Similarity 33.3%; Pred. No. 4.1;  
 Matches 12; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 7 RGRROYDHEIMEALAMVSGKMSVSKAAGITGVP 42  
 II : : : : : : : : : : : : : : : :  
 DB 24 RGSVMTLGGEAROKVAVIPSGVGDALGVGVP 59

RESULT 7  
 CIST\_SYNY3 STANDARD; PRT; 397 AA.  
 AC 059977;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Citrate synthase (EC 2.3.3.1).  
 GN GLTA OR SLT0401.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 648 to 928 of the genome."  
 RL DNA Res. 2:153-166(1995).  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + oxaloacetate -> citrate +  
 CC CoA.  
 CC -1- PATHWAY: Tricarboxylic acid cycle.  
 CC -1- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS  
 CC CAPABLE OF OXIDATIVE METABOLISM.  
 CC -1- SIMILARITY: Belongs to the citrate synthase family.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D64001: BAA10262.1: -  
 DR PIR: S74344: S74344.  
 DR HSSP: Q35354: 1AU8.  
 DR InterPro: IPR002020: Citrate\_synth.  
 DR Pfam: PF00285: citrate\_synth.1.  
 DR PRINTS: PR00143: CITRISNTFAS.  
 DR PROSITE: PS00480: CITRATE\_SYNTHASE.1.  
 KW Transferrase; Tricarboxylic acid cycle; Complete proteome.  
 FT ACT\_SITE 266 266 BY SIMILARITY.  
 FT ACT\_SITE 320 320 BY SIMILARITY.  
 SQ SEQUENCE 397 AA; 44830 MW; CD0A4A41F617F27E CRC64;

Query Match 21.2%; Score 58; DB 1; Length 397;  
 Best Local Similarity 35.0%; Pred. No. 4.9;  
 Matches 14; Conservative 7; Mismatches 15; Indels 4; Gaps 1;

QY 10 YROYDHEIMEALAMVSGKMSVSKAAGITGVPHSTLEK 49  
 I : : : : : : : : : : : : : : : :  
 DB 3 YMMDNVEFKEGLAGVPAKSRVSHVDGTGDI----LEVR 38

RESULT 8  
 RS10\_THEME STANDARD; PRT; 102 AA.  
 ID RS10\_THEME  
 AC P38518;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S10.  
 GN RPSJ OR TM1501.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=MSB8 / DSM 3109;  
 RC MEDLINE=95095941; PubMed=8002596;  
 RX MEDLINE=95095941; PubMed=8002596;  
 RA Sanangelantoni A.M., Bocchetta M., Cammarano P., Tiboni O.;  
 RT "Phylogenetic depth of S10 and spc operators: cloning and sequencing of  
 RT a ribosomal protein gene cluster from the extremely thermophilic  
 RT bacterium Thermotoga maritima."  
 RL J. Bacteriol. 176:7703-7710(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=MSB8 / DSM 3109;  
 RC MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt W.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima."  
 RL Nature 399:323-329(1999).  
 CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S10P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z21677: CAA7976.1: -  
 DR EMBL: AE001798: AAD36567.1: -  
 DR PIR: S40187: S40187.  
 DR TIGR: TM1501: -  
 DR HAMAP: MF\_00508: 1.  
 DR InterPro: IPR001848: Ribosomal\_S10.  
 DR InterPro: IPR005731: S10\_Dact.  
 DR Pfam: PF00338: Ribosomal\_S10; 1.  
 DR PRINTS: PR00971: RIBOSOMALS10.  
 DR ProDom: PD001272: Ribosomal\_S10; 1.  
 DR DR PROSITE: PS001049: rpsj\_dact; 1.  
 DR PROSITE: PS00361: RIBOSOMAL\_S10; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 102 AA; 11637 MW; ACB4D266F7D7C7F7 CRC64;

Query Match 20.9%; Score 57; DB 1; Length 102;  
 Best Local Similarity 26.8%; Pred. No. 1.4;

Matches 15; Conservative 14; Mismatches 17; Indels 10; Gaps 2;  
QY 4 RKKRGROYDHEIMEAIAVMWS-GKMSVSKAOG-----IYGVPHSTLEK 49  
ID RS10\_LISMO STANDARD; PRT; 364 AA.  
AC 026352;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan-tRNA ligase)  
GN (TPRS).  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bacterette-Stamm R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Olu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McQuigley S., Shimer G., Goyal A., Pletrovski S., Church G.M.,  
RA Daniels C.J., Mo J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
delta: functional analysis and comparative genomics.";  
RT J. Bacteriol. 179:7135-7155(1997).  
RL J. CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000812; AAB84757.1; -  
DR PIR; E69131; E69131.  
DR HAMAP; MF\_00140; -; 1.  
DR InterPro; IPR002305; tRNA-synt\_1b.  
DR InterPro; IPR004112; tRNA-synt\_1.  
DR InterPro; IPR002306; Trp\_tRNA-synt\_1b.  
DR Pfam; PF00579; tRNA-synt\_1b; 1.  
DR PRINTS; PR01039; TRNASTNTTRP.  
DR TIGRFAMs; TIGR00233; tPRS; 1.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; FALSE\_NEG.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KM Complete proteome.  
FT SITE 70 "HIGH" REGION.  
FT SITE 251 "KMSK" REGION.  
SQ SEQUENCE 364 AA; 41301 MW; C2F348903338F61D CRC64;

Query Match 20.7%; Score 56.5; DB 1; Length 364;  
Best Local Similarity 36.1%; Pred. No. 7.1;  
Matches 13; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

QY 10 YROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHST 45  
ID RS10\_LISMO STANDARD; PRT; 364 AA.  
AC 026352;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan-tRNA ligase)  
GN (TPRS).  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bacterette-Stamm R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Olu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McQuigley S., Shimer G., Goyal A., Pletrovski S., Church G.M.,  
RA Daniels C.J., Mo J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
delta: functional analysis and comparative genomics.";  
RT J. Bacteriol. 179:7135-7155(1997).  
RL J. CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000812; AAB84757.1; -  
DR PIR; E69131; E69131.  
DR HAMAP; MF\_00140; -; 1.  
DR InterPro; IPR002305; tRNA-synt\_1b.  
DR InterPro; IPR004112; tRNA-synt\_1.  
DR InterPro; IPR002306; Trp\_tRNA-synt\_1b.  
DR Pfam; PF00579; tRNA-synt\_1b; 1.  
DR PRINTS; PR01039; TRNASTNTTRP.  
DR TIGRFAMs; TIGR00233; tPRS; 1.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; FALSE\_NEG.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KM Complete proteome.  
FT SITE 70 "HIGH" REGION.  
FT SITE 251 "KMSK" REGION.  
SQ SEQUENCE 364 AA; 41301 MW; C2F348903338F61D CRC64;

QY 10 YROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHST 45  
ID RS10\_LISMO STANDARD; PRT; 364 AA.  
AC 026352;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan-tRNA ligase)  
GN (TPRS).  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bacterette-Stamm R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Olu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McQuigley S., Shimer G., Goyal A., Pletrovski S., Church G.M.,  
RA Daniels C.J., Mo J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
delta: functional analysis and comparative genomics.";  
RT J. Bacteriol. 179:7135-7155(1997).  
RL J. CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000812; AAB84757.1; -  
DR PIR; E69131; E69131.  
DR HAMAP; MF\_00140; -; 1.  
DR InterPro; IPR002305; tRNA-synt\_1b.  
DR InterPro; IPR004112; tRNA-synt\_1.  
DR InterPro; IPR002306; Trp\_tRNA-synt\_1b.  
DR Pfam; PF00579; tRNA-synt\_1b; 1.  
DR PRINTS; PR01039; TRNASTNTTRP.  
DR TIGRFAMs; TIGR00233; tPRS; 1.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; FALSE\_NEG.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KM Complete proteome.  
FT SITE 70 "HIGH" REGION.  
FT SITE 251 "KMSK" REGION.  
SQ SEQUENCE 364 AA; 41301 MW; C2F348903338F61D CRC64;

RESULT 10  
ID RS10\_LISMO STANDARD; PRT; 102 AA.  
AC 09CDM1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S10.  
GN RPSJ OR IL2100.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1350;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403.";  
RL Genome Res. 11:731-753(2001).  
RL  
CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (By  
CC similarity).  
CC  
CC -1- SIMILARITY: BELONGS TO THE S10P FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE006438; AAK06198.1; -  
DR PIR; D86887; D86887.  
DR HAMAP; MF\_00508; -; 1.  
DR InterPro; IPR001848; Ribosomal\_S10.  
DR InterPro; IPR005731; S10\_bact.  
DR Pfam; PF00338; Ribosomal\_S10; 1.  
DR PRINTS; PR00971; RIBOSOMALS10.  
DR Prodom; PD001272; Ribosomal\_S10; 1.  
DR TIGRFAMs; TIGR01049; rpsJ\_bact; 1.  
DR PROSITE; PS00361; RIBOSOMAL\_S10; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 102 AA; 11742 MW; 15D8A683A2B682A0 CRC64;

Query Match 20.5%; Score 56; DB 1; Length 102;  
Best Local Similarity 25.0%; Pred. No. 1.9;  
Matches 14; Conservative 13; Mismatches 19; Indels 10; Gaps 1;

QY 4 RKKRGROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHSTLEK 49  
ID RS10\_LISMO STANDARD; PRT; 102 AA.  
AC 0927K6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S10.  
GN RPSJ OR LMO2633 OR LIN2782.  
OS Listeria monocytogenes, and  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639; 1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21537279; PubMed=11679669;

QY 4 RKKRGROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHSTLEK 49  
ID RS10\_LISMO STANDARD; PRT; 102 AA.  
AC 0927K6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S10.  
GN RPSJ OR LMO2633 OR LIN2782.  
OS Listeria monocytogenes, and  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639; 1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21537279; PubMed=11679669;

```
RA glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Dommann E., Dominguez-Bernal G., Duchaud E., Durat L., Dussuret O.,
RA Ertlan K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Knappat G.,
RA Madueno E., Maitounan A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.,
RA Science 294:849-852(2001).
CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE S10P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AL596173; CAC98008.1; -
CC DR EMBL: AL591983; CAD00711.1; -
CC DR PIR: AH1779; AH1779.
CC DR PIR: AH1403; AH1403.
CC DR Listlist: LM002633; -
CC DR Listlist: L102782; -
CC DR HAMAP: MF_00508; -; 1.
CC DR InterPro: IPR001848; Ribosomal_S10.
CC DR InterPro: IPR005731; S10_bact.
CC DR Pfam: PF00338; Ribosomal_S10.1.
CC DR PRINTS: PR00971; RIBOSOMAL_S10.
CC DR PRODOM: PD001272; Ribosomal_S10.1.
CC DR TIGRFAMs: TIGR01049; rpsJ_bact; 1.
CC DR PROSITE: PS00361; RIBOSOMAL_S10; 1.
CC KW Ribosomal protein; Complete proteome.
CC SO SEQUENCE 102 AA; 11682 MW; 7E4B186EA533CCDB CRC64;

Query Match 20.5%; Score 56; DB 1; Length 102;
Best Local Similarity 25.0%; Pred. No. 1.9;
Matches 14; Conservative 13; Mismatches 19; Indels 10; Gaps 1;

OY 4 KKKRGROYDHEIMEAIVM-----MSGKMSVSKAAGIYGVPHSTLEYK 49
DB 4 OKIRIKAVDHRILDSAEKIVETAKRSGASVSGPIPLTEKSIYVLRVHKYK 59

RESULT 12
RS10_HELPY STANDARD; PRT; 104 AA.
AC P56017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S10.
GN rpsJ OR HP1320 OR JHP1240.
OS Helicobacter pylori (Campylobacter pylori), and
OS Helicobacter pylori j99 (Campylobacter pylori j99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
ON NCBI_TaxID=210, 85963;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
```

```
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watney L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RT Nature 388:539-547(1997).
RA [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RT Nature 397:176-180(1999).
CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE S10P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF000633; AAD08359.1; -
CC DR EMBL: AE001547; AAD06791.1; -
CC DR PIR: H64684; H64684.
CC DR TIGR: HP1320; -; 1.
CC DR HAMAP: MF_00508; -; 1.
CC DR InterPro: IPR001848; Ribosomal_S10.
CC DR InterPro: IPR005731; S10_bact.
CC DR Pfam: PF00338; Ribosomal_S10.1.
CC DR PRINTS: PR00971; RIBOSOMAL_S10.
CC DR PRODOM: PD001272; Ribosomal_S10.1.
CC DR TIGRFAMs: TIGR01049; rpsJ_bact; 1.
CC DR PROSITE: PS00361; RIBOSOMAL_S10; 1.
CC KW Ribosomal protein; Complete proteome.
CC SO SEQUENCE 104 AA; 11918 MW; 7EEA60A965B41CE1 CRC64;

Query Match 20.3%; Score 55.5; DB 1; Length 104;
Best Local Similarity 29.8%; Pred. No. 2.2;
Matches 14; Conservative 12; Mismatches 20; Indels 1; Gaps 1;

OY 5 KKKRGROYDHEIMEA-IVMWSGKMSVSKAAGIYGVPHSTLEYK 50
DB 3 KIRLIKAVDHRVLDRSVAIVAEVAKRSGSEIRGPIPLTKKRYTV 49

RESULT 13
Y769_METUA STANDARD; PRT; 169 AA.
AC O58179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M0769.
GN M0769.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
```

```

RA      Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.C., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Keriavaga A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA      Overbeek R., Klotzness E.F., Weissflog K.G., Metrick J.M., Glodek A.,
RA      Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA      Uitterlbeck T.R., Kelley J.S.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
RA      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RT      Science 273:1058-1073(1996).
RL      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U67522; AAB98774.1; -.
DR      PIR; A64396; A64396.
DR      TIGR; MJ0769; -.
DR      InterPro; IPR000792; HTH_LuxR.
DR      Hypothetical protein; Complete proteome.
RW      SEQUENCE 169 AA; 20167 MW; C1DEDFFBIFF123898 CRC64;
OY      Query Match 20.3%; Score 55.5; DB 1; Length 169;
Db      Best Local Similarity 34.1%; Pred. NO. 3.9;
Matches 14; Conservative 10; Mismatches 16; Indels 1; Gaps 1;
OY      6 KRGRROYDHEIMEAIAVMGSKMSYKAQIGYVPHSTL 46
Db      61 KRGRKKRYDDETVKKRIHELLEG-YSVAEINIGIGKSTV 100
      ||||| : : : : : : : : : : : : : : : : : : : :
RESULT 14
Y416.CHITR
ID Y416.CHITR STANDARD; PRT; 236 AA.
AC 084421;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable metal transport system ATP-binding protein CT416.
GN CT416.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_taxid=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/TW-3/Cx;
RX MEDLINE=98000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RT Science 282:754-759(1998).
RL -----
CC -!- FUNCTION: PART OF AN ATP-DRIVEN TRANSPORT SYSTEM CT415/CT416/CT417
CC FOR A METAL. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE
CC TRANSPORT SYSTEM.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001315; AAC68013.1; -.

```

DR PIR: C71517; C71517.  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR004339; ABC\_transporter.  
DR Pfam: PF00005; ABC\_tran; 1.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
KW Hypothetical protein; Transport; Inner membrane; ATP-binding;  
FT Complete proteome.  
KM NP\_BIND  
RN BIND 39  
SQ SEQUENCE 236 AA; 26193 MW; DE/4774507950365 CRC64;

Query Match Best Local Similarity 20.1%; Score 55; DB 1; Length 236;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps

OY 8 GRYYRDHDEIMEAIVM V 25  
||| | ||| ||| :  
Db 109 GRYYQEDHEAEAEALLTV 126

RESULT 15  
STH\_STAM STANDARD: PRG; 420 AA.

ID SYR\_STAM  
AC O32422;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Histidyl-tRNA synthetase (EC 6.1.1.21) (histidine-tRNA ligase)  
DE (HISRS)  
GN HIS OR SAVI631 OR SA1457 OR MM1581.  
OS Staphylococcus aureus (strain Mu50 / ATCC 706699),  
OS staphylococcus aureus (strain N315),  
OS Staphylococcus aureus (strain MW2), and  
OS Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878, 136879, 196620, 1280;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-Mu50 / ATCC 700699, and N315;  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohia T., Uchiyama I., Baba T.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumura H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mitutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kalto C.,  
RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshida C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiratsuku K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus".  
RL Lancet 357:1225-1240(2001).  
RN [2]

RN SEQUENCE FROM N.A.  
RP STRAIN-MW2;  
RX MEDLINE=22044378;  
RA Baba T., Takeuchi F., Kuroda H., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Ikama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiratsuku K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA".  
RL Lancet 359:1819-1827(2002).  
RN [3]

RN SEQUENCE FROM N.A.  
RP STRAIN-SRI17238;  
RX MEDLINE=97474251; PubMed=9335275;  
RA Fujimura T., Murakami K.;  
RT "Increase of methicillin resistance in Staphylococcus aureus caused  
RT by deletion of a gene whose product is homologous to lytic enzymes";  
RL J Bacteriol. 179:6294-6301(1997).  
RN [4]

RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=99425117; PubMed=10493797;  
RA Qiu X., Janson C.A., Blackburn M.N., Chohan I.K., Hibbs M.,

```

RA Abdel-Meguid S.S.;
RT "Cooperative structural dynamics and a novel fidelity mechanism in
RL histidyl-tRNA synthetases."
RL Biochemistry 38:12296-12304(1999).
CC -|- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +
CC diphosphate + L-histidyl-tRNA(His).
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003363; BAB57793.1; -.
DR EMBL; AP003134; BAB42723.1; -.
DR EMBL; AP004827; BAB95446.1; -.
DR EMBL; D76414; BAA23141.1; -.
DR PIR; P89945; F89945.
DR HAMAP; MF_00127; -. 1.
DR InterPro: IPR004154; HGTP_antlicodon.
DR InterPro: IPR004516; HisS
DR InterPro: IPR002314; tRNA-synt_2b.
DR InterPro: IPR006195; tRNA_antlicodon.
DR Pfam; PF03129; HGTP_antlicodon. 1.
DR Pfam; PF00587; tRNA-synt_2b; 1.
DR TIGRPFAMs; TIGR00442; hisS. 1.
DR PROSITE; PS50862; AA_tRNA_LIGASE_II; 1
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW 3D-structure; Complete proteome.
FT DISULEID 191 194
SQ SEQUENCE 420 AA; 48283 MW; 3B79521695278CA4 CRC64;

Query Match 20.1%; Score 55; DB 1; Length 420;
Best Local Similarity 29.6%; Pred. NO. 13;
Matches 16; Conservative 8; Mismatches 12; Indels 18; Gaps 2;

QY 4 RKKRGTRYROYDH-----EIMEEAIAMVMSGKMSVSKAOGIYGVPHSTL 46
DB 115 RKGKGRTRQENQGVGAIGAENPSVDAEVLAMVMTHTYQS-----FGLKHKL 161

```

Search completed: August 4, 2003, 15:05:02  
 Job time : 28 secs

THIS PAGE BLANK (USPTO)



```

A>Title: The Drosophila pipsqueak gene encodes a nuclear BRB-domain-containing protein
A:Reference number: 222927; MUID:96632300; PMID:8674425
A:Accession: T45461
A:Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: mRNA
A:Residues: 1-355; 'E', 357-1005, 'H', 1007-1020, 'Q', 1021-1061, 'ERS' <HOR>
A:Cross-References: EMBL:048358; NID:q1203906; PID:NAC47153.1; PID:q1203907
A:Experimental source: tissue type ovarian
C:Genetics:
A:Gene: pipsqueak; psq
A:Map position: II
A:Introns: 427/3
C:Function:
A:Description: required for establishing polarity of the developing egg chamber
C:Superfamily: BRCore-2 protein; POZ domain homology
F:21-123/Domain: POZ domain homology <POZ>

Query Match 32.4%; Score 88.5; DB 2; Length 1085;
Best Local Similarity 34.7%; Pred. No. 0.0062;
Matches 17; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

5 KKRGRYQYDHEIMEAIAVMYSGKMSKAGIYGVPHTLEYVKKR 53
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 773 KKEGGRKMNEDALONALEALRSQISANKASKAPGPSSTL-YKIA 820

RESULT 3
A27041
tyrosine kinase-related protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 31-Mar-1998 #sequence_revision 31-Mar-1989 #text_change 04-Feb-2000
C:Accession: A27041
R:Hallier, J.; Cole, S.; Broenner, G.; Jaeckle, H.
Genes Dev. 1, 862-867, 1987
A>Title: Dorsal and neural expression of a tyrosine kinase-related Drosophila gene during
A:Reference number: A27041; MUID:88112827; PMID:3428600
A:Accession: A27041
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-753 <HAL>
C:Genetics:
A:Gene: dTKR
A:Cross-References: FlyBase:FBgn0003715
A:Map position: 2R.60F1
A:Introns: 453/1; 497/1
C:Keywords: autophosphorylation; glycoprotein; phosphoprotein
F:9,65,187,223,224,250,611,660/Binding site: carboxydirite (Asn) (covalent) #status predicted
F:744/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 25.5%; Score 69.5; DB 2; Length 753;
Best Local Similarity 32.1%; Pred. No. 1.1;
Matches 17; Conservative 12; Mismatches 17; Indels 7; Gaps 1;

1 KPRKK-----RGRYQYDHEIMEAIAVMYSGKMSKAGIYGVPHTL 46
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 496 KMPRRKGHSAPRGGRPRSRWNTTELTEALQHVWKKMTTSGASRIFGIPVNSL 548

RESULT 4
G65007
hypothetical protein b2345 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: G65007
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.: Rose, D.U.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65007
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-352 <BLAT>

```

```

A:Cross-references: GB:AE000323; GB:000096; NID:q1788664; PIDN:AAC75405.1; PID:q1788664
A:Experimental source: strain K-12, substrain M6165
C:Superfamily: Escherichia coli hypothetical protein b2345

Query Match      22.3%; Score 61; DB 2; Length 352;
Best Local Similarity 28.8%; Pred. No. 5.8;
Matches 17; Conservative 15; Mismatches 15; Indels 12; Gaps 3;

RESULT 5
B40628
resolvasse homolog - Enterococcus faecium transposon Tn1546
C:Species: Enterococcus faecium
C:Date: 21-Sep-1993 #sequence_revision 07-Oct-1994 #text_change 21-Aug-1998
C:Accession: B40628
J.Arthur, M.; Molinas, C.; Depardieu, F.; Courvalin, P.
J. Bacteriol. 175, 117-127, 1993
A>Title: Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide
A:Reference number: A40628; MUID:93106944; PMID:8380148
A:Comments: B4147, p1P816
A:Accession: B40628
A:Molecule type: DNA
A:Residues: 1-191 <AR>
A:Cross-references: GB:M97297
A>Note: Sequence extracted from NCBI backbone (NCBIN:121238, NCBI:P.121242)
C:Superfamily: transposase repressor
C:Keywords: DNA binding

Query Match      21.6%; Score 59; DB 2; Length 191;
Best Local Similarity 29.6%; Pred. No. 5.3;
Matches 16; Conservative 13; Mismatches 23; Indels 2; Gaps 1;

QY      1 KQPKKKRRRYROY--DHEIMEEALAMVMSGMSKAGQIGYGVPHSTLEKYKVE 52
Db      135 KKEGFKGRLLKHKHKNAGMNYAVKLKREGMTVNOICETITVNSRSLYRKLSE 188

RESULT 6
S51715
hypothetical protein - Chlamydia trachomatis (fragment)
C:Species: Chlamydia trachomatis
C:Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 11-Jan-2000
C:Accession: S51715
R.Pohlner, J.
submitted to the EMBL Data Library, April 1994
A:Description: Sequence of the redA gene from Chlamydia trachomatis L2.
A:Reference number: S51714
A:Accession: S51715
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-419 <POH>
A:Cross-references: EMBL:Z32530; NID:9607060; PIDN:CA83540.1; PID:9607062
C:Superfamily: Chlamydia hypothetical protein CT648

Query Match      21.6%; Score 59; DB 2; Length 419;
Best Local Similarity 32.7%; Pred. No. 13;
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 3;

QY      1 KQPKKKGRRYROYDHEIMEEALAMVMSG-----MSVSKAQGIGYGVPH 43
Db      227 KQGLRSGKTYD-PLTKETACVWNGKGOVIYGYKATITETROYHGVPH 275

RESULT 7
F71488
hypothetical protein CT648 - Chlamydia trachomatis (serotype D, strain W3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 10-Dec-1999
C:Accession: F71488

```



```

Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: F71488
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-424 <ARN>
A:Cross-references: GB:AE001335; GB:AE001273; NID:93329091; PIDN:AA668825.1; PID:9332909
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT648
C:Superfamily: Chlamydia hypothetical protein CT648

Query Match .21.6%; Score 59; DB 2; Length 424;
Best Local Similarity 32.7%; Pred. No. 13;
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 3;

OY 1 KQPKKRGRYQYDHEIMEEALMVMSGK-----MSYSKQGYGVP 43
|| : : : || : : : || : : : || : : : || : : : ||
Db 232 KQGLRSGKY--YD-PLTKKEIACVNGKRGVYIGRAYIETROIYHGVP 280

RESULT 8
A57364
recombination protein recA - Myxococcus xanthus
N:Alternate names: recombination A
C:Species: Myxococcus xanthus
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 02-Feb-2001
C:Accession: A57364
J:Norick, N.; Hsu, M.Y.; Inouye, S.; Inouye, M.
J. Bacteriol. 177, 4179-4182, 1995
A:Title: Two recA genes in Myxococcus xanthus.
A:Reference number: A57364; MUID:95332233; PMID:7608099
A:Accession: A57364
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-342 <RES>
A:Cross-references: GB:I40367; NID:9848993; PIDN:AA37000.1; PID:9848994
C:Genetics:
A:Gene: recA
C:Superfamily: recombination protein recA
C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; nucleotide binding; P-loop;
F:68-75/Region: nucleotide-binding motif A (P-loop)
F:142-147/Region: nucleotide-binding motif B
F:74/Binding site: ATP (Lys) #status predicted

Query Match 21.2%; Score 58; DB 2; Length 342;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 12; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

OY 7 RGRYQYDHEIMEEALMVMSGKMSVSKAOGIYGP 42
|| : : : || : : : || : : : || : : : || : : : ||
Db 24 RGSVMTLGFAREQKAVIPSGSYVDRLALGVGYP 59

RESULT 9
S74344
citrate synthase gltA - Synechocystis sp. (strain PCC 6803)
N:Alternate names: hypothetical protein sl10401
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74344
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Matsubae, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74344
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```

A:	Residues:	1-97	<KAM>
A:	Cross-references:	EMBL:D64001; GB:AB001339; NID:g1001102; PIDN:BAAI0262.1; PID:g100	
A:	Note:	the nucleotide sequence was submitted to the EMBL Data Library, June 1996	
C:	Genetics:		
A:	Gene:	glcA	
A:	Start codon:	GTC	
C:	Superfamily:	citrinate (sl)-synthase	
Query Match	21.2%; Score 58; DB 2; Length 397;		
Best Local Similarity	35.0%; Pred. No. 16;		
Matches	14; Conservative 7; Mismatches 15; Indels 4; Gaps 1;		
Oy	10 YRQYDHEIMEEALIAMVMSGKMYSKAOGITYGVPHSTLETK 49		
Dn	3 YMTNDNEVFKEGLAGVPAAKSRSVSHVDGTGI---LEYR 38		
RESULT 10			
T18531	Tractin - medicinal leech		
C:	Species:	Hirudo medicinalis (medicinal leech)	
C:	Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000	
C:	Accession:	T18531	
R:	Name:	J. Jellies, J.; Johansen, K.M.; Johansen, J.	
J.	Cell Biol.	138, 143-157, 1997	
A:	Title:	Differential glycosylation of Tractin and leechCAM, two novel Ig-superfamily	
A:	Reference number:	Z18951; MURID:97362067; PMID:9214388	
A:	Accession:	T18531	
A:	Status:	Preliminary: translated from GR/EMBL/DDBJ	
A:	Molecule type:	mRNA	
A:	Residues:	1-1880 <HUA>	
A:	Cross-references:	EMBL:U92813; NID:g2275259; PID:g2275260; PIDN:AAC47654.1	
Query Match	21.2%; Score 58; DB 2; Length 1880;		
Best Local Similarity	29.4%; Pred. No. 92;		
Matches	20; Conservative 11; Mismatches 19; Indels 18; Gaps 3;		
Oy	4 RKKRGTRYOYDH-----ETMEEALIAMVMSG--KMSVS KAOGIY--GVPHST 45		
Dn	889 RKRRSRSGKNCHGDENVIGHVTNLMPDSMVAKAAVLVLNGCKKGELSEPTITFYPEVGVRSKV 948		
Oy	46 LEYVKER 53		
Dn	949 MEPRVMER 956		
RESULT 11			
H75179	Hypothetical protein PAB2413 - Pyrococcus abyssi (strain Orsay)		
C:	Species:	Pyrococcus abyssi	
C:	Date:	20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000	
C:	Accession:	H75179	
R:	Anonymous,	Genoscope	
	submitted to the EMBL Data Library, July 1999		
A:	Description:	Pyrococcus abyssi genome sequence: insights into archaeal chromosome s	
A:	Reference number:	A75001	
A:	Accession:	H75179	
A:	Status:	Preliminary	
A:	Molecule type:	DNA	
A:	Residues:	1-139 <KAM>	
A:	Cross-references:	GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49519.1; PID:g545	
A:	Experimental source:	strain Orsay	
C:	Genetics:		
A:	Gene:	PAB2413	
C:	Superfamily:	Pyrococcus abyssi hypothetical protein PAB2413	
Query Match	21.1%; Score 57.5; DB 2; Length 139;		
Best Local Similarity	30.6%; Pred. No. 5.8;		
Matches	15; Conservative 12; Mismatches 21; Indels 1; Gaps 1;		
Oy	5 KKRRRYOYDHEIMEEALIAMVMSGKMYSKAOGITYGVPHSTLETKVER 53		
Dn	85 KGGRRPKRYSYKTLMIEELLKGGTPREISRD-LGPRTLYLLAKR 132		



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2003, 14:54:34 ; Search time 83 seconds

(without alignments)  
101.356 Million cell updates/sec

Title: US-10-016-768a-2

Perfect score: 273  
Sequence: 1 KOPRRKRGRRQYDHEIMEE.....KAQGYVPHSTLEKYKER 53

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_19Jun03.\*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273	100.0	53	AAE24592	Human E93 program
2	273	100.0	442	AAE24371	Human E93 program
3	273	100.0	630	ABG17942	Novel human diagno
4	234	100.7	104	ABP32451	Human ORF1424 prot
5	233.5	85.5	54	AAE24593	Fish E93 programme
6	229	83.9	53	AAE24594	Mouse E93 programme
7	165	60.4	53	AAE24370	Fruit fly E93 prog
8	165	60.4	1140	ABR71145	Drosophila melanog
9	165	60.4	1165	AAE24372	Fruit fly E93 prog

10	163	59.7	53	AAE24595	Nematode E93 progr
11	88.5	32.4	1046	ABE67028	Drosophila melanog
12	88.5	32.4	1064	ABE59068	Drosophila melanog
13	70.5	25.8	661	ABE53113	Drosophila melanog
14	69.5	25.5	1046	ABE59307	Drosophila melanog
15	63	23.1	457	ABE68768	Drosophila melanog
16	62	22.7	325	ABE66413	Thermus thermophil
17	62	22.3	352	AAE98384	Escherichia coli p
18	61	22.3	494	ABP26391	Streptococcus poly
19	61	22.3	494	ABP29724	Streptococcus poly
20	59.5	21.8	263	ABE64196	Drosophila melanog
21	59	21.6	191	AAE24294	Resolvase encoded
22	59	21.6	434	AAE37533	Chlamydia trachoma
23	59	21.6	484	ABE68769	Drosophila melanog
24	58	21.2	141	AAO05999	Human polypeptide
25	58	21.2	364	AAW34643	Thermococcus alpha
26	58	21.2	364	ABE28281	T. alcaliphilus AE
27	58	21.2	364	ABG70786	T. alcaliphilus AE
28	58	21.2	364	ABG70828	T. alcaliphilus AE
29	58	21.2	364	ABG70919	T. alcaliphilus al
30	58	21.2	364	ABG31855	Thermococcus alca
31	58	21.2	364	ABG74442	Thermococcus alca
32	58	21.2	364	ABG73912	T. alcaliphilus al
33	58	21.2	364	ABG73965	Thermococcus alca
34	57.5	21.1	957	ABE64362	Drosophila melanog
35	57	20.9	298	AAE07753	An oocyte maturati
36	56.5	20.7	2027	AAE10987	Mouse presentin/C
37	56	20.5	102	ABE55454	Lactococcus lactis
38	56	20.5	102	ABE49361	Listeria monocytog
39	56	20.5	1693	AAE80141	Human protein SEQ
40	56	20.5	1723	AAW79157	Human protein SEQ
41	55.5	20.3	104	AAU35850	Helicobacter pylor
42	55.5	20.3	104	AAU36004	Helicobacter pylor
43	55.5	20.3	828	ABE5774	Drosophila melanog
44	55	20.1	204	AAE91636	C glutamicum prote
45	55	20.1	236	AAE37820	Amino acid sequenc

#### ALIGNMENTS

RESULT 1	AAE24592	AAE24592 standard; Protein: 53 AA.
ID	AAE24592	
AAE24592:		
04-OCT-2002 (first entry)		
XX	Human E93 programmed cell death modulating protein conserved domain.	
XX	Human; cancer: programmed cell death modulating protein; adenocarcinoma;	
AC	cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;	
XX	neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;	
DT	Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;	
XX	aplastic anaemia; ischemic injury; myocardial infarction; stroke;	
DE	reperfusion injury; toxin-induced disease; genetic immunodeficiency;	
XX	vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;	
XX	KW myeloma; nocrotropic; vasotropic; immunostimulant; cerebroprotective;	
XX	KW cardiant; E93 protein.	
OS	Homo sapiens.	
PN	WO200234882-A2.	
PD	02-MAY-2002.	
XX	29-OCT-2001; 2001WO-US48053.	
PF	27-OCT-2000; 2000US-243865P.	
XX	(UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.	
XX		

PI Baehrecke EH;  
XX  
XX WPI: 2002-479717/51.  
DR  
XX  
XX  
PT Novel programmed cell death modulating proteins, useful for treating or  
PT preventing disorders associated with abnormal cell proliferation and  
PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
PI infarction  
XX  
XX  
PS Claim 1; Fig 1; 88pp: English.  
XX  
XX The present invention relates to novel programmed cell death modulating  
CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase  
CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences  
CC of the invention are used as vaccines and in gene therapy. The present  
CC sequence is human E93 programmed cell death modulating protein conserved  
CC domain.  
XX  
XX Sequence 53 AA:  
SQ  
Query Match 100.0%; Score 273; DB 23; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2, 5e-33;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KOPRRKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVER 53  
DB 1 KOPRRKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVER 53  
RESULT 2  
AAE24371  
ID AAE24371 standard; Protein: 442 AA.  
XX  
AC AAE24371;  
XX  
DT 04-OCT-2002 (first entry)  
XX  
DE Human E93 programmed cell death modulating protein.  
XX  
XX Human; cancer: programmed cell death modulating protein; adenocarcinoma;  
KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;  
KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;  
KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
KW myeloma; neurotropic; vasotropic; immunostimulant; cerebroprotective;  
KW cardiac; E93 protein.  
XX  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
FT Domain 353..405  
FT /note="Conserved domain"  
XX  
XX WO200234882-A2.  
XX  
XX 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-US48053.  
PF

XX  
XX 27-OCT-2000; 2000US-243865P.  
PR  
XX  
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
PA  
XX  
XX Baehrecke EH;  
XX  
XX WPI: 2002-479717/51.  
DR  
XX  
XX  
PT Novel programmed cell death modulating proteins, useful for treating or  
PT preventing disorders associated with abnormal cell proliferation and  
PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
PI infarction  
XX  
XX  
PS Claim 1; Fig 4; 88pp: English.  
XX  
XX The present invention relates to novel programmed cell death modulating  
CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase  
CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences  
CC of the invention are used as vaccines and in gene therapy. The present  
CC sequence is human E93 programmed cell death modulating protein.  
XX  
XX Sequence 442 AA:  
SQ  
Query Match 100.0%; Score 273; DB 23; Length 442;  
Best Local Similarity 100.0%; Pred. No. 4e-32;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KOPRRKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVER 53  
DB 353 KOPRRKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVER 405  
RESULT 3  
ABG17942  
ID ABG17942 standard; Protein: 630 AA.  
XX  
AC ABG17942;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17933.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW  
XX  
OS Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI

XX WPI: 2001-639362/73.  
DR N-PSDB: AAS82129.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20, SEQ ID NO 48301, 103pp: English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 630 AA:  
OY  
Query Match 100.0%; Score 273; DB 22; Length 630;  
Best Local Similarity 100.0%; Pred. No. 6,4e-32;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 KOPRRKRGRRYQDHEIMEAIVMVGSKMSVSKAAGIYGVPSHTEYKVKER 53  
541 KOPRRKRGRRYQDHEIMEAIVMVGSKMSVSKAAGIYGVPSHTEYKVKER 593  
RESULT 4  
ID ABB32451 standard; Protein: 104 AA.  
XX  
AC ABB32451:  
XX  
DT 09-JUL-2002 (first entry)  
XX  
DE Human ORF1424 protein, SEQ ID NO:2848.  
XX  
KW Human: ORF: open reading frame; ORFX: drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; haematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; tumour inhibition; bodily characteristics; fertility;  
KW behaviour; cancer; proliferative disorder; neurological disorder;  
KW cardiovascular disease; immune system disorder; organ transplantation;  
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
KW vasotropic; antiproliferative; antidiabetic; cytostatic; neutrophic;  
KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;  
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KW dermatological; analgesic; virucide; antibacterial; fungicide.  
XX  
OS Homo sapiens.  
XX  
XX MO200190366-A2.  
XX  
XX 29-NOV-2001.

XX  
XX 24-MAY-2001; 2001MO-US17076.  
PF  
XX  
XX 24-MAY-2000; 2000US-206690P.  
PR  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Leach MD, Shinkets RA:  
PI  
XX  
XX WPI: 2002-106200/14.  
DR  
DR N-PSDB: ABB76477.  
XX  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation  
XX  
PS Claim 10; Page 971-972; 2508pp: English.  
XX  
XX Sequences ABB31028-ABB35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABB75054-  
CC ABB75587 represent cDNAs encoding them. The invention also encompasses  
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or  
CC activity, and methods of screening individuals for a predisposition to an  
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
CC range of biological activities, such as cytokine, cell proliferation,  
CC cell differentiation, immune modulation, haematopoiesis regulation,  
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
CC and antimicrobial activity, and may also be involved in the determination  
CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
CC nucleic acids and antibodies may be used in the treatment of cancers,  
CC other proliferative disorders such as psoriasis and benign tumours,  
CC neurological disorders such as epilepsy and Alzheimer's disease,  
CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes. In the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX  
CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
XX  
SQ Sequence 104 AA:  
OY  
Query Match 85.7%; Score 234; DB 23; Length 104;  
Best Local Similarity 84.9%; Pred. No. 4,4e-27;  
Matches 45; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
Db 1 KOPRRKRGRRYQDHEIMEAIVMVGSKMSVSKAAGIYGVPSHTEYKVKER 53  
8 KOPRRKRGRRYQDHEIMEAIVMVGSKMSVSKAAGIYGVPSHTEYKVKER 60  
RESULT 5  
ID AAE24593 standard; Protein: 54 AA.  
XX  
XX AAE24593:  
XX  
XX 04-OCT-2002 (first entry)  
XX

DE Fish E93 programmed cell death modulating protein conserved domain.  
XX  
XX Fish; cancer; programmed cell death modulating protein; adenocarcinoma;  
KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
KW aplastic anaemia; ischemic injury; myocardial infarction; stroke;  
KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;  
KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
KW myeloma; neotropic; vasotropic; immunostimulant; cerebroprotective;  
KW cardiact; E93 protein.  
XX  
XX Tetraodon nigroviridis.  
OS  
XX WO200234882-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-US48053.  
PF  
XX 27-OCT-2000; 2000US-243865P.  
PR  
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
PA  
XX Baehrecke EH;  
PI  
XX WPI; 2002-479717/51.  
DR  
XX  
XX Novel programmed cell death modulating proteins, useful for treating or  
PT preventing disorders associated with abnormal cell proliferation and  
PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
PT infarction  
XX  
XX Claim 1; Fig 1; 88pp; English.  
PS  
XX The present invention relates to novel programmed cell death modulating  
CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase  
CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences  
CC of the invention are used as vaccines and in gene therapy. The present  
CC sequence is fish E93 programmed cell death modulating protein conserved  
CC domain.  
XX  
XX Sequence 54 AA:  
SQ  
XX  
XX Query Match 85.5%; Score 233.5; DB 23; Length 54;  
Best Local Similarity 81.5%; Pred. No. 2.2e-27;  
Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

OY 1 KOPRRKRGROYDHEIMEEA-1AMVSGKMSVSKAOGIYGVPHSTLEKYKER 53  
DB 1 KOPRRKRGROYDHEIMEEA-1AMVSGKMSVSKAOGIYGVPHSTLEKYKER 54

RESULT 6  
AAE24594  
ID AAE24594 standard; Protein; 53 AA.  
XX  
XX AAE24594;  
XX  
XX 04-OCT-2002 (first entry)

XX  
XX Mouse E93 programmed cell death modulating protein conserved domain.  
DE  
XX  
XX Mouse; cancer; programmed cell death modulating protein; adenocarcinoma;  
KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
KW aplastic anaemia; ischemic injury; myocardial infarction; stroke;  
KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;  
KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
KW myeloma; neotropic; vasotropic; immunostimulant; cerebroprotective;  
KW cardiact; E93 protein.  
XX  
XX Mus musculus.  
OS  
XX WO200234882-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-US48053.  
PF  
XX 27-OCT-2000; 2000US-243865P.  
PR  
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
PA  
XX Baehrecke EH;  
PI  
XX WPI; 2002-479717/51.  
DR  
XX  
XX Novel programmed cell death modulating proteins, useful for treating or  
PT preventing disorders associated with abnormal cell proliferation and  
PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
PT infarction  
XX  
XX Claim 1; Fig 1; 88pp; English.  
PS  
XX The present invention relates to novel programmed cell death modulating  
CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase  
CC in cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences  
CC of the invention are used as vaccines and in gene therapy. The present  
CC sequence is mouse E93 programmed cell death modulating protein conserved  
CC domain.  
XX  
XX Sequence 53 AA:  
SQ  
XX  
XX Query Match 83.9%; Score 229; DB 23; Length 53;  
Best Local Similarity 81.1%; Pred. No. 1e-26;  
Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 KOPRRKRGROYDHEIMEEA-1AMVSGKMSVSKAOGIYGVPHSTLEKYKER 53  
DB 1 KOPRRKRGROYDHEIMEEA-1AMVSGKMSVSKAOGIYGVPHSTLEKYKER 53

RESULT 7  
AAE24370  
ID AAE24370 standard; Protein; 53 AA.  
XX  
XX AAE24370;  
XX  
XX

[illegible]







XX	27-SEP-2001.
PD	
XX	
XX	23-MAR-2001; 2001WO-USO9231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI: 2001-656860/75.
DR	N-PSDB; ABL11131.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Disclosure; SEQ ID NO 27876; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABLJ30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XQ	Sequence 1046 AA;

```

Query Match          32.4%; Score 88.5; DB 22; Length 1046;
Best Local Similarity 34.7%; Pred. No. 0.0007;
Matches 17; Conservative 14; Mismatches 17; Indels 1; Gaps 1
OY      5 KKRGRYQYDHEIMEALAMVMSGKMSYSKAQGIYGVHSTLEYKVER 53
      11 - - - - - : : : : : : : : : : : : : : : : : : :
Db      755 KKEGCTKSMNEDALONALELRSGQISANKKSKAFGIPSSYL-YKIA 802

RESULT 12
ID      ABB59068 standard; Protein; 1064 AA.
XX      ABB59068;
XX      ABB59068;
AC      ABB59068;
DT      26-MAR-2002 (first entry)
XX
DE      Drosophila melanogaster polypeptide SEQ ID NO 3996.
XX
KW      Drosophila; developmental biology; cell signalling; insecticide;
KW      pharmaceutical.
XX
XX      Drosophila melanogaster.
OS      Drosophila melanogaster.
PN      WO200171042-A2.
PD      27-SEP-2001.
XX
PF      23-MAR-2001; 2001WO-US09231.
XX
PR      23-MAR-2000; 2000US-191637P.
PR      11-JUL-2000; 2000US-0614150.
XX
PA      (PEKE ) PE CORP NY.
XX
PI      Venter JC, Adams M, Li PWD, Myers EW;
XX      WPI: 2001-656860/75.
XX

```

DR N-PSDB: ABL03171.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS  
PS Disclosure: SEQ ID NO 3996; 21np + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.

	Query Match	32.4%	Score 88.5	DB 22:	Length 1064;	
	Best Local Similarity	34.7%	Pred. NO. 0.00072;			
Matches	17: Conservative	14: Mismatches	17: Indels	1: Gaps	1:	
Dc	773 KKEGGSNMEDALQNLRLRSGISANKSKAFGISSTL-VRIARR	820				
Oy	5 KKRGROYDHEIMEEAIAMVSCGMYSKAGIYGVHSTLEYVKER	53				
	: ::::   :   ::  :	:				
RESULT 13						
ID	ABB63113 standard; Protein; 661 AA.					
XX	ABB63113;					
AC	ABBB3113;					
XX	26-MAR-2002 (first entry)					
DT						
XX	Drosophila melanogaster polypeptide SEQ ID NO 16131.					
DE						
XX	Drosophila; developmental biology; cell signalling; insecticide;					
KW	pharmaceutical.					
OS	Drosophila melanogaster.					
XX	WO200171042-A2.					
PN						
XX	27-SEP-2001.					
PD						
XX	23-MAR-2001; 2001WO-USO9231.					
PF						
PR	23-MAR-2000; 2000US-191637P.					
PP	11-JUL-2000; 2000US-0614150.					
XA	(PEKE ) PE CORP NY.					
PA						
PI	Venter JC, Adams M, Li PWD, Myers EW;					
DR	WPI: 2001-6556860/75.					
N-PSDB;	ABL07216.					
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -					
PS	Disclosure; SEQ ID NO 16131; 21pp + Sequence Listing; English.					
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention					



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2003, 15:01:44 : Search time 95 Seconds  
(without alignments)  
143.966 Million cell updates/sec

Title: US-10-016-768a-2

Perfect score: 273  
Sequence: 1 KQPRKRRGRYQYDHEIMEAIAVMVSGKMSVSKAGIYGVPHSTLEYVKER 53

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP viirus:\*  
16: SP bacteriaph:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the distribution, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	273	100.0	213	4	Q96NKL1
2	273	100.0	517	11	Q8CJG4
3	242	88.6	396	11	Q8C9Q0
4	242	88.6	433	11	Q8BGT2
5	242	88.6	572	4	Q8GJN0
6	242	88.6	619	4	Q8N3I6
7	165	60.4	1165	5	Q9VD60
8	165	60.4	1398	5	Q9VDM8
9	163	59.7	185	5	Q22051
10	163	59.7	185	5	Q22051
11	88.5	32.4	660	5	Q24457
12	88.5	32.4	1064	5	Q9VSN1
13	88.5	32.4	1085	5	Q24455
14	78.5	28.8	652	5	Q77168
15	72	26.4	107	12	Q91IHO
16	70.5	25.8	661	5	Q9V8S2

17	69.5	25.5	1046	5	Q9W0W2	Q9W0W2 drosophila
18	69	25.3	365	3	Q9P3F3	Q9P3F3 neurospora
19	67	24.5	924	16	Q8CJL0	Q8CJL0 streptomyces
20	64	23.4	393	11	Q8C9J6	Q8C9J6 mus musculus
21	63	23.1	977	5	Q9W0K7	Q9W0K7 drosophila
22	63	23.1	977	5	Q9UJH7	Q9UJH7 drosophila
23	61	22.3	478	16	Q8R5T4	Q8R5T4 thermomana
24	61	22.3	494	16	Q8R5H9	Q8R5H9 streptococcus
25	61	22.3	444	16	Q8D2T2	Q8D2T2 streptococcus
26	60.5	22.2	325	3	Q9UVG7	Q9UVG7 magnaporthe
27	59.5	21.8	267	5	Q9GP27	Q9GP27 drosophila
28	59.5	21.8	267	5	Q9V5X7	Q9V5X7 beta vulgar
29	59.5	21.8	551	8	Q9MF64	Q9MF64 enterococcus
30	59	21.6	176	2	Q8KSEA	Q8KSEA xanthomonas
31	59	21.6	308	16	Q8PCX8	Q8PCX8 leptospira
32	59	21.6	342	16	Q8F5P1	Q8F5P1 chlamydia t
33	59	21.6	419	2	Q46445	Q46445 chlamydia t
34	59	21.6	424	16	Q84654	Q84654 anquilla ja
35	59	21.6	470	13	Q8Q6W1	Q8Q6W1 drosophila
36	59	21.6	1067	5	Q8VYH3	Q8VYH3 staphylococcus
37	58.5	21.4	224	2	Q8VYQ3	Q8VYQ3 staphylococcus
38	58.5	21.4	224	2	Q8GFL0	Q8GFL0 staphylococcus
39	58.5	21.4	224	2	Q8G8W8	Q8G8W8 staphylococcus
40	58	21.2	193	4	Q8N9Y1	Q8N9Y1 homo sapien
41	58	21.2	271	17	Q8RMT7	Q8RMT7 methanococcus
42	58	21.2	306	16	Q8PG12	Q8PG12 xanthomonas
43	58	21.2	413	4	Q8IWF9	Q8IWF9 homo sapien
44	58	21.2	761	16	Q8DIA7	Q8DIA7 synecococcus
45	58	21.2	1880	5	Q18465	Q18465 hirudo med

## ALIGNMENTS

RESULT 1  
Q96NKL1 PRELIMINARY: PRT: 213 AA.

AC Q96NKL1;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ30696.  
OC Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP TISSUE=Brain;  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakatsuki A., Ishii S., Yamamoto J., Isono Y., Kawai-Hito Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Magatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;  
RT "NEO human cDNA sequencing project."  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK055258; BAB70892.1; -  
KW Hypothetical protein  
SQ SEQUENCE 213 AA: 23477 MW: 4D7F6CABF95251B2 CRC64;

Query Match 100.0%; Score 273; DB 4; Length 213;  
Best Local Similarity 100.0%; Pred. No. 2.4e-26;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KQPRKRRGRYQYDHEIMEAIAVMVSGKMSVSKAGIYGVPHSTLEYVKER 53  
|||||  
DB 124 KQPRKRRGRYQYDHEIMEAIAVMVSGKMSVSKAGIYGVPHSTLEYVKER 176  
|||||

RESULT 2

```
Q8CJG4 ID 08CJG4 PRELIMINARY; PRT; 517 AA.
AC 08CJG4
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Transcription factor MLR1.
GN MLR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA Kunieda T., Park J., Takeuchi H., Kubo T.;
RT "Mus musculus mlr1 and mlr2 mRNA for transcription factor MLR1 and
RT MLR2."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076078; BAC20954.1; -
SQ SEQUENCE 517 AA; 57316 MW; C97403D3D296C52E CRC64;

Query Match 100.0%; Score 273; DB 11; Length 517;
Best Local Similarity 100.0%; Pred. No. 6.4e-26;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 53
Db 429 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 481

RESULT 3
Q8C900 ID 08C900 PRELIMINARY; PRT; 396 AA.
AC 08C900:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK041621; BAC31007.1; -
KW Hypothetical protein.
FT NON_TER 396
SQ SEQUENCE 396 AA; 43085 MW; EE4A585FE2336E35 CRC64;

Query Match 88.6%; Score 242; DB 11; Length 396;
Best Local Similarity 86.8%; Pred. No. 3.8e-22;
Matches 46; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 53
Db 337 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 389

RESULT 4
Q8BGT2 ID 08BGT2 PRELIMINARY; PRT; 433 AA.
AC 08BGT2:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
```

```
DE Transcription factor MLR2 (Hypothetical protein).
GN MLR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA Kunieda T., Park J., Takeuchi H., Kubo T.;
RT "Mus musculus mlr1 and mlr2 mRNA for transcription factor MLR1 and
RT MLR2."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AB076079; BAC20955.1; -
DR EMBL; AK041090; BAC30816.1; -
KW Hypothetical protein.
SQ SEQUENCE 433 AA; 47124 MW; 736656D1F7E9A041 CRC64;

Query Match 88.6%; Score 242; DB 11; Length 433;
Best Local Similarity 86.8%; Pred. No. 4.2e-22;
Matches 46; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 53
Db 337 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 389

RESULT 5
Q96JN0 ID 096JN0 PRELIMINARY; PRT; 572 AA.
AC 096JN0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein KIAA1795 (Fragment).
GN KIAA1795.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama W., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058698; BAB47424.1; -
KW Hypothetical protein.
FT NON_TER 572
SQ SEQUENCE 572 AA; 62730 MW; FB0A401D3F060DF4 CRC64;

Query Match 88.6%; Score 242; DB 4; Length 572;
Best Local Similarity 86.8%; Pred. No. 5.8e-22;
Matches 46; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 53
Db 476 KOPRKKRGYROYDHEIMEAIAVMGSKMSVSKAGIYGVPHSTLEYKVKR 528

RESULT 6
```

08N3L6 ID 08N3L6 PRELIMINARY; PRT; 619 AA.  
 AC 08N3L6;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DE 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 GN Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wambutt R., Heubner D., Mewes H.W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL834245; CAD38921.1; -  
 KM Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 619 AA; 67378 MW; 791286C6F8A5110 CRC64;  
 Query Match 88.6%; Score 242; DB 4; Length 619;  
 Best Local Similarity 86.8%; Pred. No. 6.3e-22;  
 Matches 46; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 KOPRRKRGRRYDHEIMEALAMVMSGMSVSKAGIYGVPHSTLEKVKER 53  
 DB 523 KOPRRKRGRRYDHEIMEALAMVMSGMSVSKAGIYGVPHSTLEKVKER 575  
 RESULT 7  
 O9VD60 PRELIMINARY; PRT; 1165 AA.  
 AC 09VD60;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 GN CG18389 protein.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein J., Brotter P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein J., Bolshakov S.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,  
 RA Finkler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,  
 RA Jaiswal N., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jaiswal B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kromoller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminber J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kromoller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnitsk F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003737; AAF55940.3; -  
 DE FlyBase: FBgn0013948; EIP93F.  
 SQ SEQUENCE 1165 AA; 123976 MW; A2556014070BED8D CRC64;  
 Query Match 60.4%; Score 165; DB 5; Length 1165;  
 Best Local Similarity 60.4%; Pred. No. 6.5e-12;  
 Matches 32; Conservative 6; Mismatches 15; Indels 0; Gaps 0;  
 OY 1 KOPRRKRGRRYDHEIMEALAMVMSGMSVSKAGIYGVPHSTLEKVKER 53  
 DB 758 KOPRRKRGRRYDHEIMEALAMVMSGMSVSKAGIYGVPHSTLEKVKER 810  
 RESULT 8  
 O95YM8 PRELIMINARY; PRT; 1598 AA.  
 AC 095YM8;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DE 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 GN MblK-1 protein.  
 OS Apis mellifera (Honeybee).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OX Ephydriidae; Drosophilidae; Drosophila.  
 RN NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dublin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
 RA Jatali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshirefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palzozolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weislock G.M., Weissbach J.,  
 RA Williams S.M., Woodde T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri L.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibeagwam C., Jatali M., Kruse D., Li P., Mattei B., Moshirefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Pfeiffer B., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
 RA Phounenavong S., Pittman G.S., Puri V., Richards S., Schaefer F.,  
 RA Stapleton M., Strong R., Svirskis R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bereman C., Bernan B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnicki F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [4]

```

RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003829; AAM68770.1; -.
DR FlyBase: FBgn0004399; psq.
DR InterPro: IPR002197; HTH_Fis.
DR TIGRfams: TIGR01199; HTH_Fis; 1.
SQ SEQUENCE 645 AA; 70298 MW; 4B72F47175060529 CRC64;

Query Match 32.4%; Score 88.5; DB 5; Length 645;
Best Local Similarity 34.7%; Pred. No. 0.015;
Matches 17; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

OY 5 KKRGRYROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHSTLEYKVKER 53
DB 348 KKEGKTSWNEALQNALALRSQISANKSKAFGIPSTL-YK1ARR 401

RESULT 11
ID 024457 PRELIMINARY; PRT; 660 AA.
AC 024457;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 02, Last annotation update)
DE PIPISQUEAK protein (ORF-B);
GN PSQ OR CG3368.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phyllozoa; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-96134923; PubMed-8557044;
RA Weber U., Siegel V., Mlodzik M.;
RT "Pipisquak encodes a novel nuclear protein required downstream of
RT seven-up for the development of photoreceptors R3 and R4.";
RL EMBL: X90986; CAA62475.1; -.
DR FlyBase: FBgn0004399; psq.
DR InterPro: IPR002197; HTH_Fis.
DR TIGRfams: TIGR01199; HTH_Fis; 2.
SQ SEQUENCE 660 AA; 71818 MW; 6E251F440326547F CRC64;

Query Match 32.4%; Score 88.5; DB 5; Length 660;
Best Local Similarity 34.7%; Pred. No. 0.015;
Matches 17; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

OY 5 KKRGRYROYDHEIMEAIAVMWSGKMSVSKAOGIYGVPHSTLEYKVKER 53
DB 348 KKEGKTSWNEALQNALALRSQISANKSKAFGIPSTL-YK1ARR 395

RESULT 12
ID 09V5N1 PRELIMINARY; PRT; 1064 AA.
AC 09V5N1; Q9V5N2; Q24184; Q24187;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Psg protein (LD33470p);
GN PSQ OR CG3368.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydrozoa; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE-Ovary;
RX MEDLINE-95220671; PubMed-7705633;
RA Horowitz H., Berg C.A.;
RT "Aberrant splicing and transcription termination caused by P element
RT insertion into the intron of a Drosophila gene.";
RL Genetics 139:327-335(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE-Ovary;
RX MEDLINE-96232300; PubMed-8674425;
RA Horowitz H., Berg C.A.;
RT "The Drosophila pipsquak gene encodes a nuclear BTB-domain-containing
RT protein required early in oogenesis.";
RL Development 122:1859-1871(1996).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS A AND 2).
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Fostier C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodex A., Gong F., Garg N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Iyegyan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Del Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sriden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svartkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Zaveri T., Morley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zeng J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter C.J.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A/1 (SHOWN HERE), B AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL: U48358; AAC47153.1; -.
DR EMBL: U48402; AAC47154.1; -.
DR EMBL: AE003829; AAF58769.1; -.
DR EMBL: AE003829; AAF58770.1; -.
DR EMBL: AY069588; AAL39733.1; -.

```









THIS PAGE BLANK (USPTO)

Mon Aug 4 16:28:23 2003

us-10-016-768a-2.rapb

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2003, 15:06:54 ; Search time 50 Seconds

(without alignments)  
125.885 Million cell updates/sec

Title: US-10-016-768a-2

Perfect score: 273

Sequence: 1 KQPKRRGRYROYDHEIMEE.....KAQGYVPHSTLEYKVKR 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/2/pubppaa/PCF\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/2/pubppaa/PCFUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*

12: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273	100.0	53	US-10-016-768-2	Sequence 2, Appl1
2	273	100.0	44	US-10-016-768-8	Sequence 8, Appl1
3	233.5	85.5	54	US-10-016-768-3	Sequence 3, Appl1
4	229	83.9	53	US-10-016-768-4	Sequence 4, Appl1
5	165	60.4	53	US-10-016-768-1	Sequence 10, Appl1
6	165	60.4	1165	US-10-016-768-10	Sequence 10, Appl1
7	163	59.7	53	US-10-016-768-5	Sequence 5, Appl1
8	66	24.2	378	US-10-156-761-12022	Sequence 12022, A
9	62	22.3	325	US-09-938-901-2	Sequence 2, Appl1
10	61	22.3	352	US-09-741-669-432	Sequence 432, Appl1
11	58	21.2	364	US-09-886-400-4	Sequence 4, Appl1
12	58	21.2	364	US-10-112-357-4	Sequence 4, Appl1
13	58	21.2	364	US-10-114-403-4	Sequence 4, Appl1
14	58	21.2	364	US-10-116-606-4	Sequence 4, Appl1
15	58	21.2	364	US-10-112-331-4	Sequence 4, Appl1

16	58	21.2	364	US-10-112-377-4	Sequence 2, Appl1
17	58	21.2	364	US-10-116-581-4	Sequence 11443, A
18	58	21.2	364	US-10-112-442-4	Sequence 11597, A
19	58	21.2	364	US-10-112-418-4	Sequence 5390, Ap
20	58	21.2	364	US-10-114-083-4	Sequence 5343, Ap
21	56.5	20.7	2027	US-10-234-961-2	Sequence 12646, A
22	55.5	20.3	104	US-09-815-242-11443	Sequence 12817, A
23	55.5	20.3	104	US-09-815-242-11597	Sequence 13723, A
24	55	20.1	204	US-09-738-626-5390	Sequence 54, Appl1
25	55	20.1	420	US-09-815-242-5343	Sequence 35, Appl1
26	55	20.1	420	US-09-815-242-12648	Sequence 12459, A
27	54.5	20.1	420	US-09-815-242-12817	Sequence 9784, Ap
28	54.5	20.0	470	US-09-815-242-13723	Sequence 10607, A
29	53.5	19.6	384	US-09-825-414-54	Sequence 13196, A
30	53.5	19.6	1597	US-09-832-292-35	Sequence 38, Appl1
31	53	19.4	102	US-10-156-761-12459	Sequence 38, Appl1
32	53	19.4	102	US-10-156-761-9784	Sequence 2977, Ap
33	52.5	19.2	433	US-09-815-242-11597	Sequence 4, Appl1
34	52	19.0	102	US-09-815-242-13196	Sequence 3025, Ap
35	52	19.0	102	US-09-925-637-38	Sequence 8025, Ap
36	52	19.0	102	US-09-908-931B-34	Sequence 18, Appl1
37	52	19.0	102	US-10-084-205-38	Sequence 162, Appl1
38	52	19.0	129	US-09-764-891-2977	Sequence 22, Appl1
39	52	19.0	552	US-09-817-764-4	
40	52	19.0	724	US-10-128-714-3025	
41	52	19.0	1139	US-10-128-714-8025	
42	51.5	18.9	349	US-09-742-153-18	
43	51.5	18.9	358	US-10-278-536-162	
44	51.5	18.9	534	US-09-742-153-22	
45	51.5	18.9	1399	US-09-742-153-14	

#### ALIGNMENTS

RESULT 1

US-10-016-768-2

Sequence 2, Application US/10016768

Publication No. US2002012443A1

GENERAL INFORMATION:

APPLICANT: Baehrcke, Eric H.

TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH

FILE REFERENCE: 4115-131

CURRENT APPLICATION NUMBER: US/10/016, 768

CURRENT FILING DATE: 2001-10-29

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 53

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MISC\_FEATURE

LOCATION: (1)..(54)

OTHER INFORMATION: X CAN BE ANY AMINO ACID

US-10-016-768-2

Query Match

Best Local Similarity 100.0%; Score 273; DB 14; Length 53;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db

1 KQPKRRGRYROYDHEIMEEALIAVWMSGKSVSKAAGIYVPHSTLEYKVKR 53

US-10-016-768-8

Sequence 8, Application US/10016768

Publication No. US2002014243A1

GENERAL INFORMATION:

APPLICANT: Baehrcke, Eric H.

TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH

FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8  
LENGTH: 442  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-016-768-8

Query Match 100.0%; Score 273; DB 14; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1,3e-30;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53  
Db 353 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 405

## RESULT 3

US-10-016-768-3  
Sequence 3, Application US/10016768  
Publication No. US20020142443A1  
GENERAL INFORMATION:  
APPLICANT: Baehrcke, Eric H.  
TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 54  
TYPE: PRT  
ORGANISM: T. nigroviridis  
US-10-016-768-3

Query Match 85.5%; Score 233.5; DB 14; Length 54;  
Best Local Similarity 81.5%; Pred. No. 1,3e-25;  
Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

OY 1 KOPRKKGRYQYDHEIMEA-IAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53  
Db 1 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 54

## RESULT 4

US-10-016-768-4  
Sequence 4, Application US/10016768  
Publication No. US20020142443A1  
GENERAL INFORMATION:  
APPLICANT: Baehrcke, Eric H.  
TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 53  
TYPE: PRT  
ORGANISM: M. musculus  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (1)..(54)  
OTHER INFORMATION: X can be any amino acid  
US-10-016-768-4

Query Match 83.9%; Score 229; DB 14; Length 53;  
Best Local Similarity 81.1%; Pred. No. 5,6e-25;  
Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53  
Db 1 KHPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53

## RESULT 5

US-10-016-768-1  
Sequence 1, Application US/10016768  
Publication No. US20020142443A1  
GENERAL INFORMATION:  
APPLICANT: Baehrcke, Eric H.  
TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (1)..(54)  
OTHER INFORMATION: X can be any amino acid  
US-10-016-768-1

Query Match 60.4%; Score 165; DB 14; Length 53;  
Best Local Similarity 60.4%; Pred. No. 5,6e-16;  
Matches 32; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

OY 1 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53  
Db 1 KGTTPKRGKYRNDSDSLVEAVKAVQVREMSVHRAGSYGVPHSTLEYKVKER 53

## RESULT 6

US-10-016-768-10  
Sequence 10, Application US/10016768  
Publication No. US20020142443A1  
GENERAL INFORMATION:  
APPLICANT: Baehrcke, Eric H.  
TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 10  
LENGTH: 1165  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-016-768-10

Query Match 60.4%; Score 165; DB 14; Length 1165;  
Best Local Similarity 60.4%; Pred. No. 2,7e-16;  
Matches 32; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

OY 1 KOPRKKGRYQYDHEIMEAIAVMGSKMSVSKAAGIYGVPHSTLEYKVKER 53  
Db 758 KGTTPKRGKYRNDSDSLVEAVKAVQVREMSVHRAGSYGVPHSTLEYKVKER 810

## RESULT 7

US-10-016-768-5  
Sequence 5, Application US/10016768  
Publication No. US20020142443A1  
GENERAL INFORMATION:  
APPLICANT: Baehrcke, Eric H.  
TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
FILE REFERENCE: 4115-131  
CURRENT APPLICATION NUMBER: US/10/016,768  
CURRENT FILING DATE: 2001-10-29

NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (1)-(54)  
OTHER INFORMATION: X CAN BE ANY AMINO ACID  
US-10-016-768-5

Query Match 59.7%; Score 163; DB 14; Length 53;  
Best Local Similarity 56.6%; Pred. No. 1,1e-15;  
Matches 30; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 1 KQPRKRGROYDHEIMEAIAWMSGKMSVSKAGCIYVPHSTLEYKVER 53  
DB 1 KRSRPRKRGQYKDYKDNALDEAVRSVRGEMTVHRAGSFQVPHSTLEYKVER 53

RESULT 8  
US-10-156-761-12022  
Sequence 12022, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 12022  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-12022

Query Match 24.2%; Score 66; DB 15; Length 378;  
Best Local Similarity 36.4%; Pred. No. 0.55;  
Matches 12; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 3 PRKRGROYDHEIMEAIAWMSGKMSVSKA 35  
DB 58 PEKENSRYEKFDYPIKEQVANLTKGQDVAYA 90

RESULT 9  
US-09-938-901-2  
Sequence 2, Application US/09938901  
Publication No. US20030008291A1  
GENERAL INFORMATION:  
APPLICANT: Kuramitsu Seiki,  
APPLICANT: Yokoyama Shigeyuki  
TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME  
FILE REFERENCE: PH-1261-US  
CURRENT APPLICATION NUMBER: US/09/938,901  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: JP2001-47762  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2

LENGTH: 325  
TYPE: PRT  
ORGANISM: Thermus thermophilus  
US-09-938-901-2

Query Match 22.7%; Score 62; DB 11; Length 325;  
Best Local Similarity 32.6%; Pred. No. 1.7;  
Matches 15; Conservative 8; Mismatches 9; Indels 14; Gaps 2;

QY 3 PRKRGROYDHEIMEAIAWMSGKMSV-----SKAGCIYV 42  
DB 208 PRKRRAR-----EERLVALVLLGKRGVHLERLEGRFQGLYGV 245

RESULT 10  
US-09-741-669-432  
Sequence 432, Application US/09741669  
Patent No. US20020022718A1  
GENERAL INFORMATION:  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Ohlsen, Karl L.  
TITLE OF INVENTION: Genes identified as required for  
FILE REFERENCE: ELITRA.009A  
CURRENT APPLICATION NUMBER: US/09/741,669  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: US 60/173005  
PRIOR FILING DATE: 1999-12-23  
NUMBER OF SEQ ID NOS: 481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 432  
LENGTH: 352  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-741-669-432

Query Match 22.3%; Score 61; DB 9; Length 352;  
Best Local Similarity 28.8%; Pred. No. 2.5;  
Matches 17; Conservative 15; Mismatches 15; Indels 12; Gaps 3;

QY 5 KRGROYDHEI-----MEEAIAWMSGKMSVSKAGCIYVPH-STLEYKVER 53  
DB 271 KRGRARIRYDFQVDAYKODLKEIYVTLGNKINNGHSGKIT--KHFDLBEKIER 327

RESULT 11  
US-09-886-400-4  
Sequence 4, Application US/09886400  
Patent No. US20020045226A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: Murphy, Dennis  
TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE  
FILE REFERENCE: DIVER110-4  
CURRENT APPLICATION NUMBER: US/09/886,400  
CURRENT FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: 09/619,032  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/407,806  
PRIOR FILING DATE: 1999-09-20  
PRIOR APPLICATION NUMBER: 08/613,220  
PRIOR FILING DATE: 1996-03-08  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 364  
TYPE: PRT  
ORGANISM: Thermococcus alcaliphilus  
US-09-886-400-4

Query Match 21.2%; Score 58; DB 9; Length 364;

Best Local Similarity 36.4%; Pred. No. 7;  
Matches 16; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

OY 1 KOPKKRGRRYROYDHEIME--EATAMVSGMSYSKAQGIYGP 42

Db 165 KAQREKRFRYISYLLGLRELKRAIKIVFEGKVTLKAVKDIEAVP 208

## RESULT 12

US-10-112-357-4

; Sequence 4, Application US/10112357  
; Publication No. US20020115099A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: Murphy, Dennis  
; APPLICANT: Ried, John  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
; FILE REFERENCE: DIVER1120-4  
; CURRENT APPLICATION NUMBER: US/10/112,357  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: 09/886,400  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 09/619,032  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/407,806  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 08/613,220  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Thermococcus alcaliphilus  
US-10-112-357-4

Query Match 21.2%; Score 58; DB 14; Length 364;

Best Local Similarity 36.4%; Pred. No. 7;

Matches 16; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

OY 1 KOPKKRGRRYROYDHEIME--EATAMVSGMSYSKAQGIYGP 42

Db 165 KAQREKRFRYISYLLGLRELKRAIKIVFEGKVTLKAVKDIEAVP 208

## RESULT 13

US-10-114-403-4

; Sequence 4, Application US/10114403  
; Publication No. US20020115100A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: Murphy, Dennis  
; APPLICANT: Ried, John  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
; FILE REFERENCE: DIVER1120-4  
; CURRENT APPLICATION NUMBER: US/10/114,403  
; PRIOR FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: 09/886,400  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 09/619,032  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/407,806  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 08/613,220  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Thermococcus alcaliphilus  
US-10-114-403-4

Query Match 21.2%; Score 58; DB 14; Length 364;

Best Local Similarity 36.4%; Pred. No. 7;  
Matches 16; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

OY 1 KOPKKRGRRYROYDHEIME--EATAMVSGMSYSKAQGIYGP 42

Db 165 KAQREKRFRYISYLLGLRELKRAIKIVFEGKVTLKAVKDIEAVP 208

## RESULT 14

US-10-116-606-4

; Sequence 4, Application US/10116606  
; Publication No. US20020119515A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: Murphy, Dennis  
; APPLICANT: Ried, John  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
; FILE REFERENCE: DIVER1120-4  
; CURRENT APPLICATION NUMBER: US/10/116,606  
; PRIOR FILING DATE: 2002-04-03  
; PRIOR APPLICATION NUMBER: 09/886,400  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 09/619,032  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/407,806  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 08/613,220  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Thermococcus alcaliphilus  
US-10-116-606-4

Query Match 21.2%; Score 58; DB 14; Length 364;

Best Local Similarity 36.4%; Pred. No. 7;

Matches 16; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

OY 1 KOPKKRGRRYROYDHEIME--EATAMVSGMSYSKAQGIYGP 42

Db 165 KAQREKRFRYISYLLGLRELKRAIKIVFEGKVTLKAVKDIEAVP 208

## RESULT 15

US-10-112-331-4

; Sequence 4, Application US/10112331  
; Publication No. US20020119550A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: Murphy, Dennis  
; APPLICANT: Ried, John  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
; FILE REFERENCE: DIVER1120-4  
; CURRENT APPLICATION NUMBER: US/10/112,331  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: 09/886,400  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 09/619,032  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/407,806  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 08/613,220  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Thermococcus alcaliphilus  
US-10-112-331-4

Mon Aug 4 16:28:23 2003

us-10-016-768a-2.iapb

Page 5

US-10-112-331-4

Query Match 21.2%; Score 58; DB 14; Length 364;

Best Local Similarity 36.4%; Pred. No. 7; Mismatches 18; Indels 2; Gaps 1;

Matches 16; Conservative 8; Mismatches 18; Indels 2; Gaps 1;

OY 1 KOPRKRGRYROYDHEIME--EAIAMVSGMSYSKAQIYGP 42

Db 165 KAQREKFRFRYISYLLGLRELKRAIKLVFECKVTLLKANVDIEAVP 208

Search completed: August 4, 2003, 15:15:50

Job time : 51 secs

THIS PAGE BLANK





# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 100337**

**TO: Minh-Tam Davis**  
**Location: CM-1/8A01/8E12**  
**Art Unit: 1642**  
**Monday, August 04, 2003**

**Case Serial Number: 10/016768**

**From: Toby Port**  
**Location: Biotech-Chem Library**  
**CM1-6A04**  
**Phone: 308-3534**

**[toby.port@uspto.gov](mailto:toby.port@uspto.gov)**

### **Search Notes**

Dear Examiner Davis,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK**